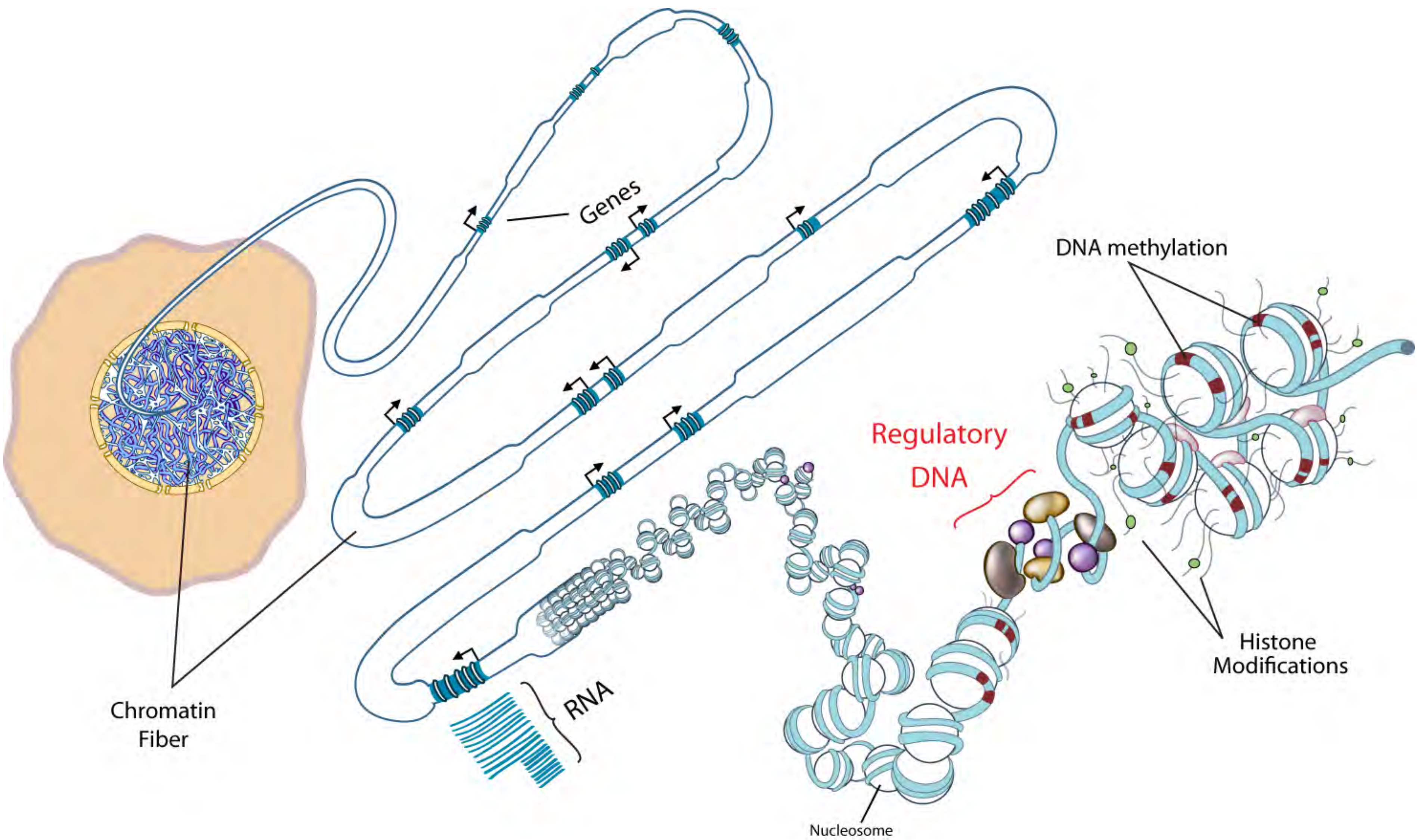
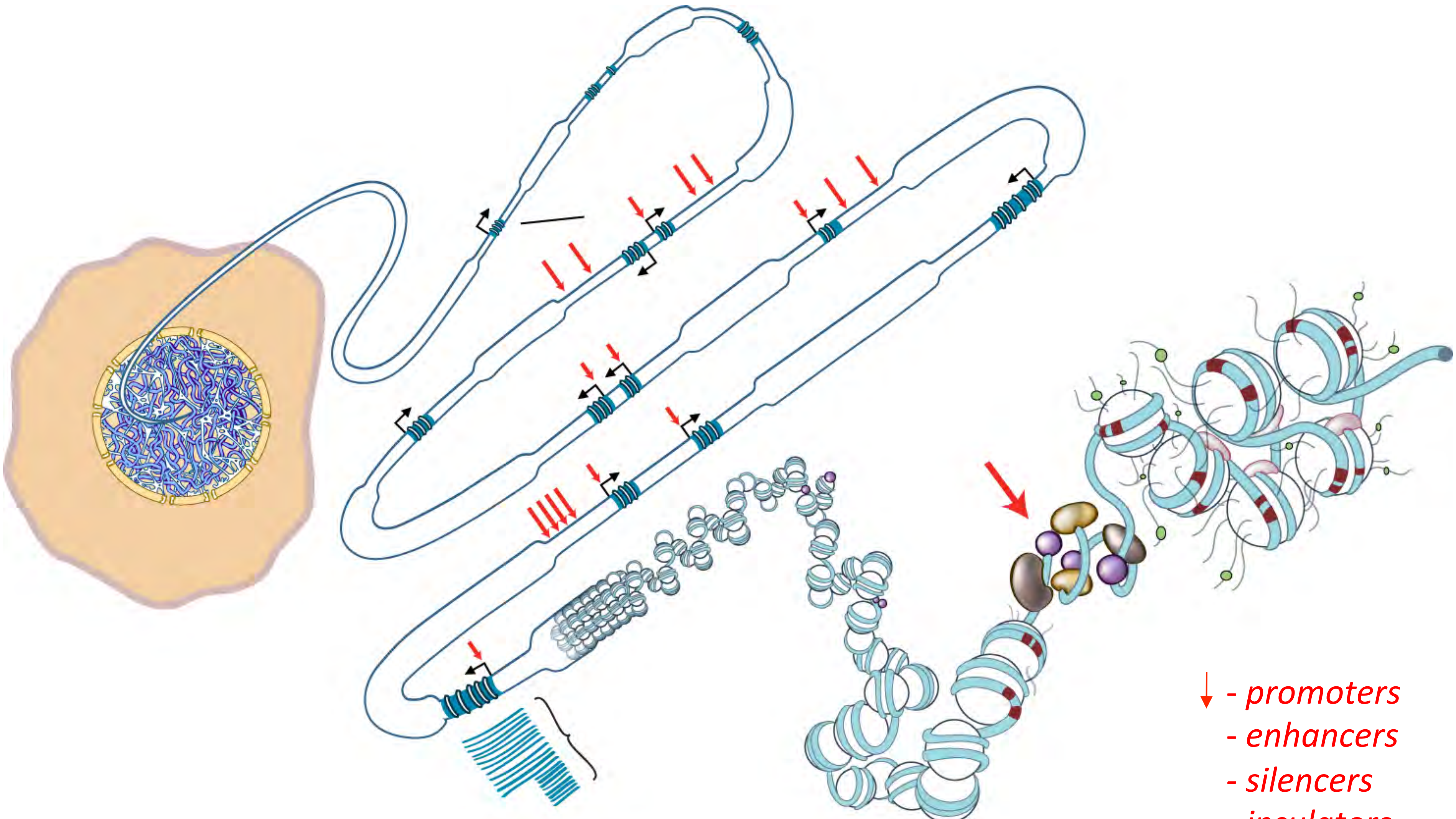
Gene regulation and common diseases and traits

John A. Stamatoyannopoulos, M.D.
Depts. of Genome Sciences & Medicine
University of Washington

The living genome

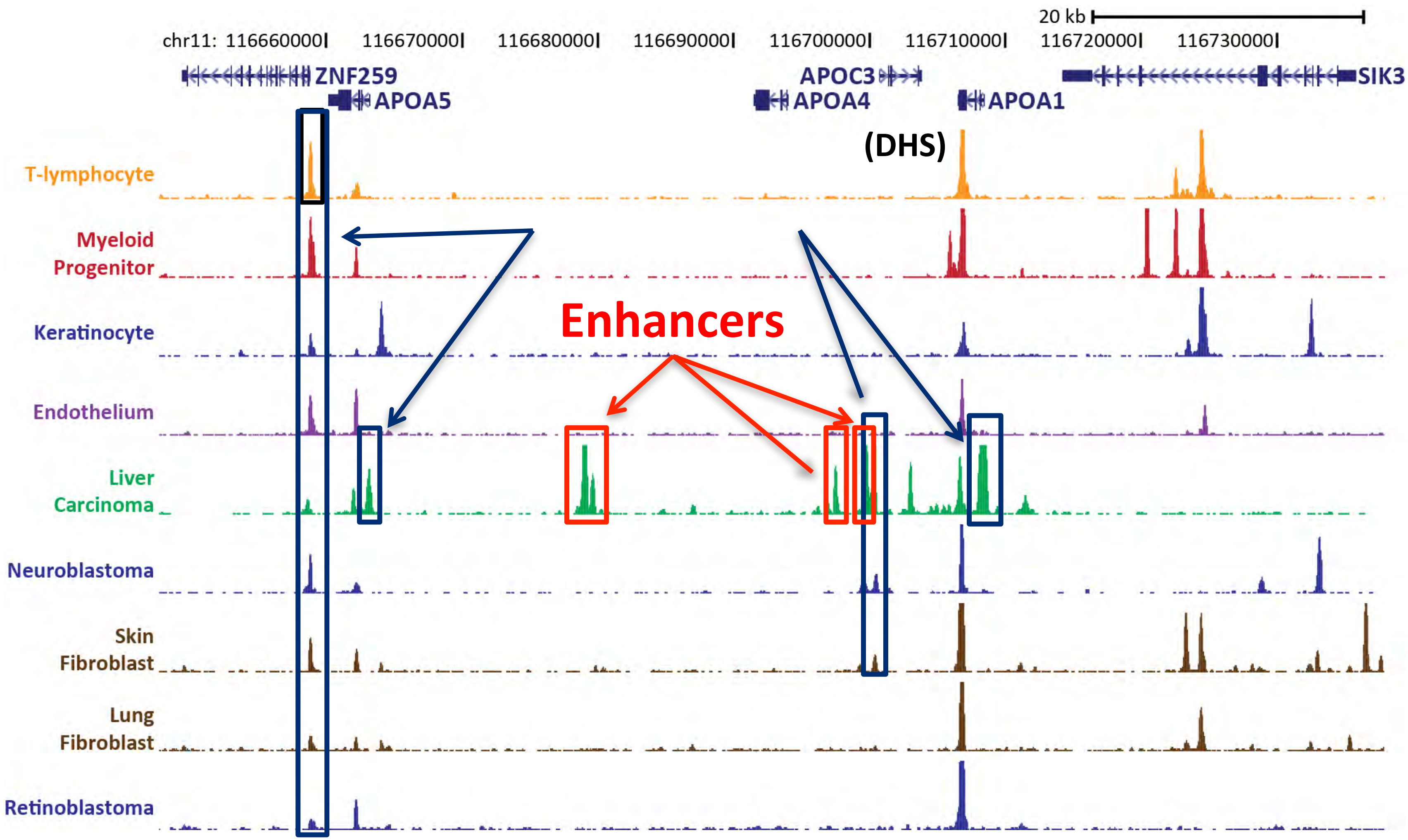


The living genome



- promoters
- enhancers
- silencers
- insulators
- etc.

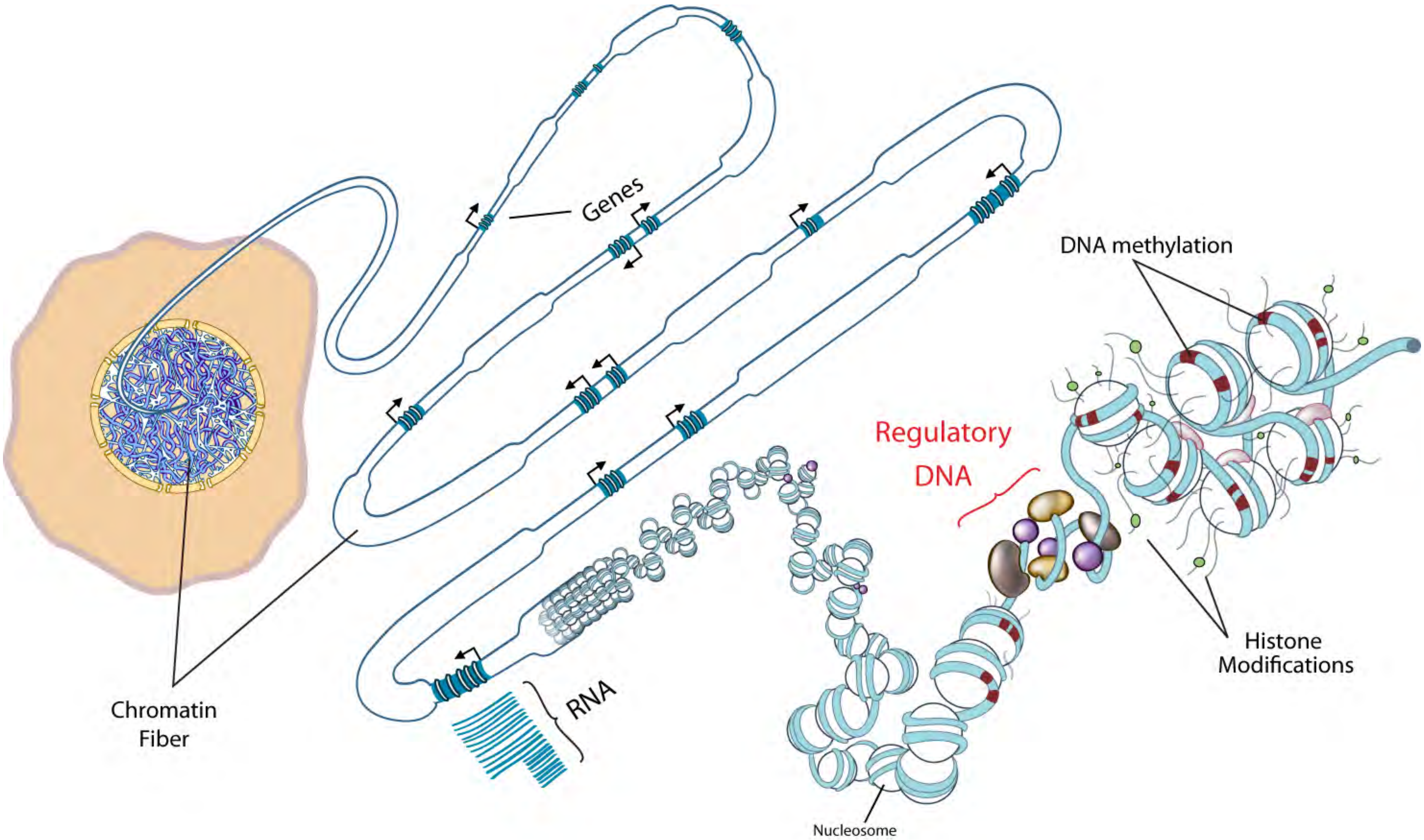
DNaseI hypersensitive sites (DHSs) mark regulatory DNA



~100,000 – 250,000 DHSs per cell type (0.5-1.5% of genome)

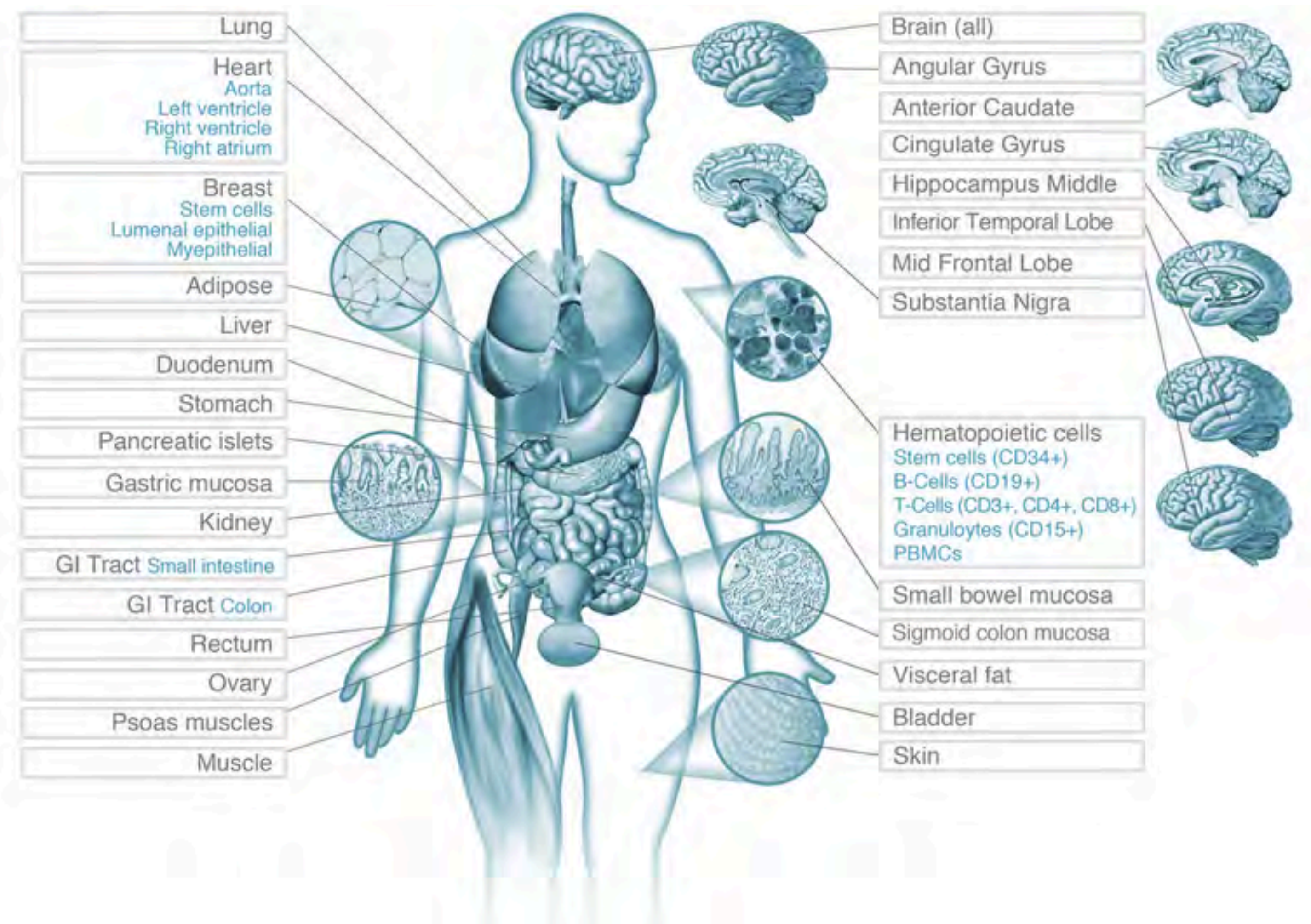
Thurman et al, Nature, 2012

The living genome



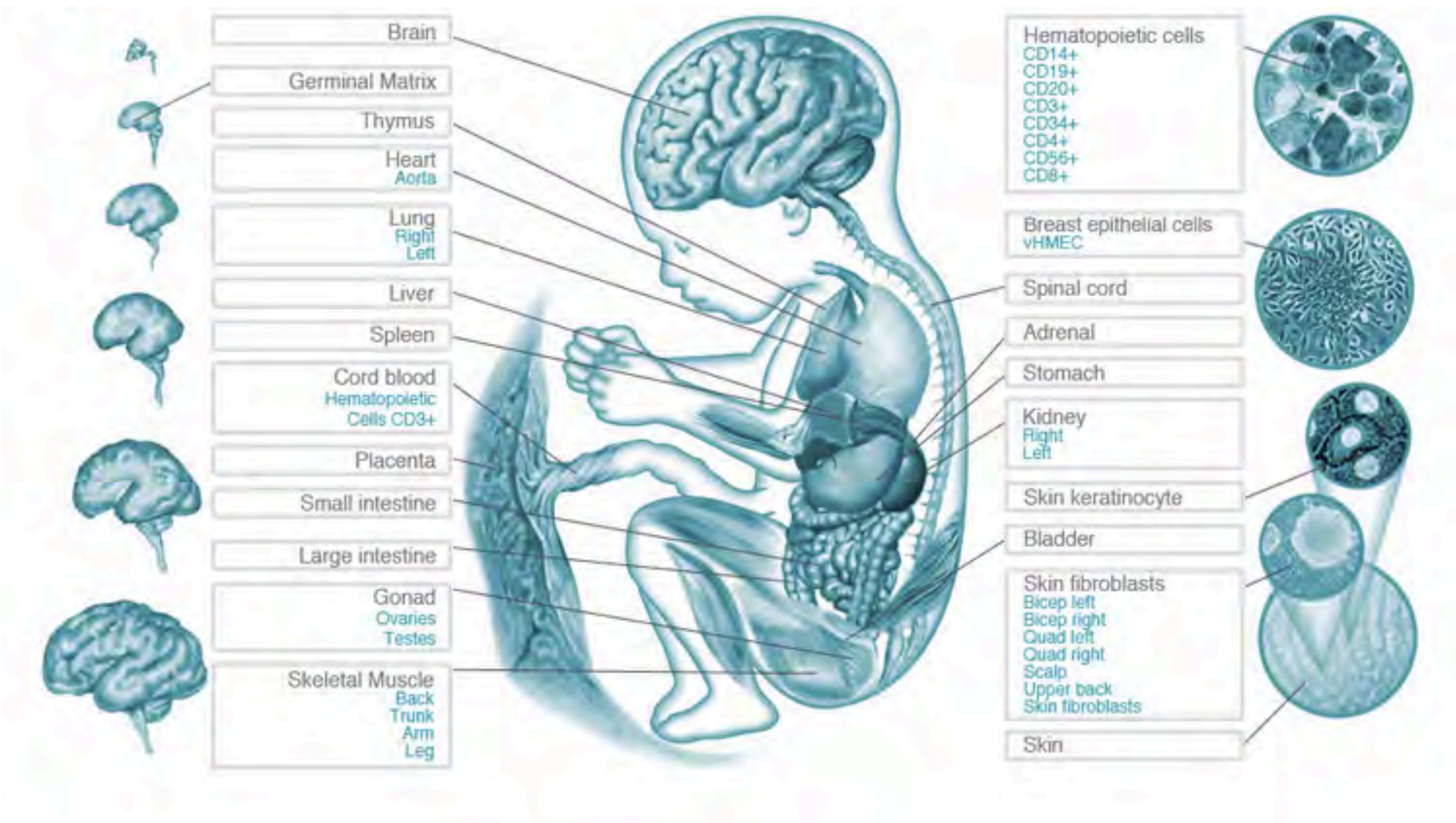
Surveying the regulatory DNA landscape

Adult cells and tissues

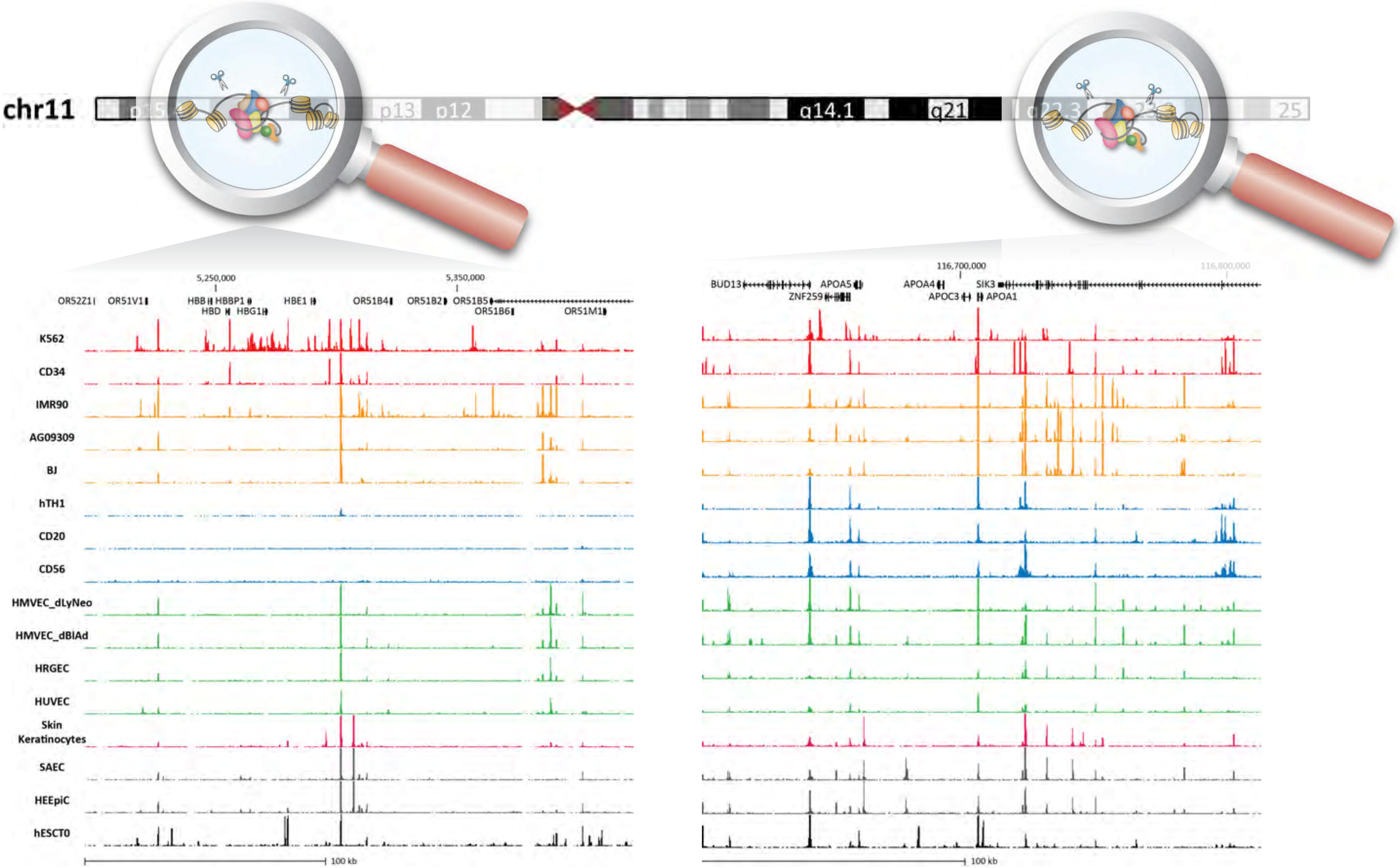


Surveying the regulatory DNA landscape

Developing cells and tissues

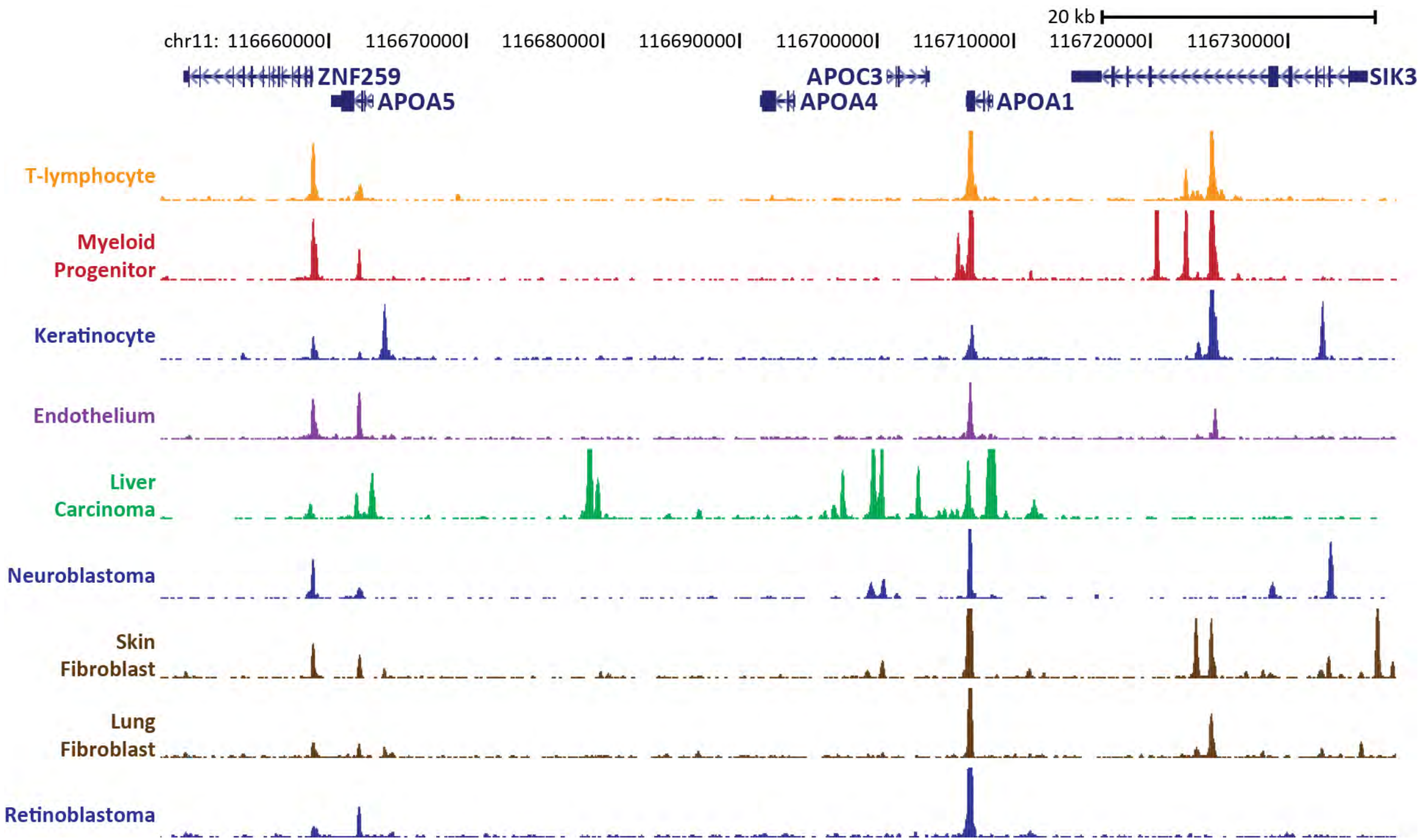


An available atlas of human regulatory DNA

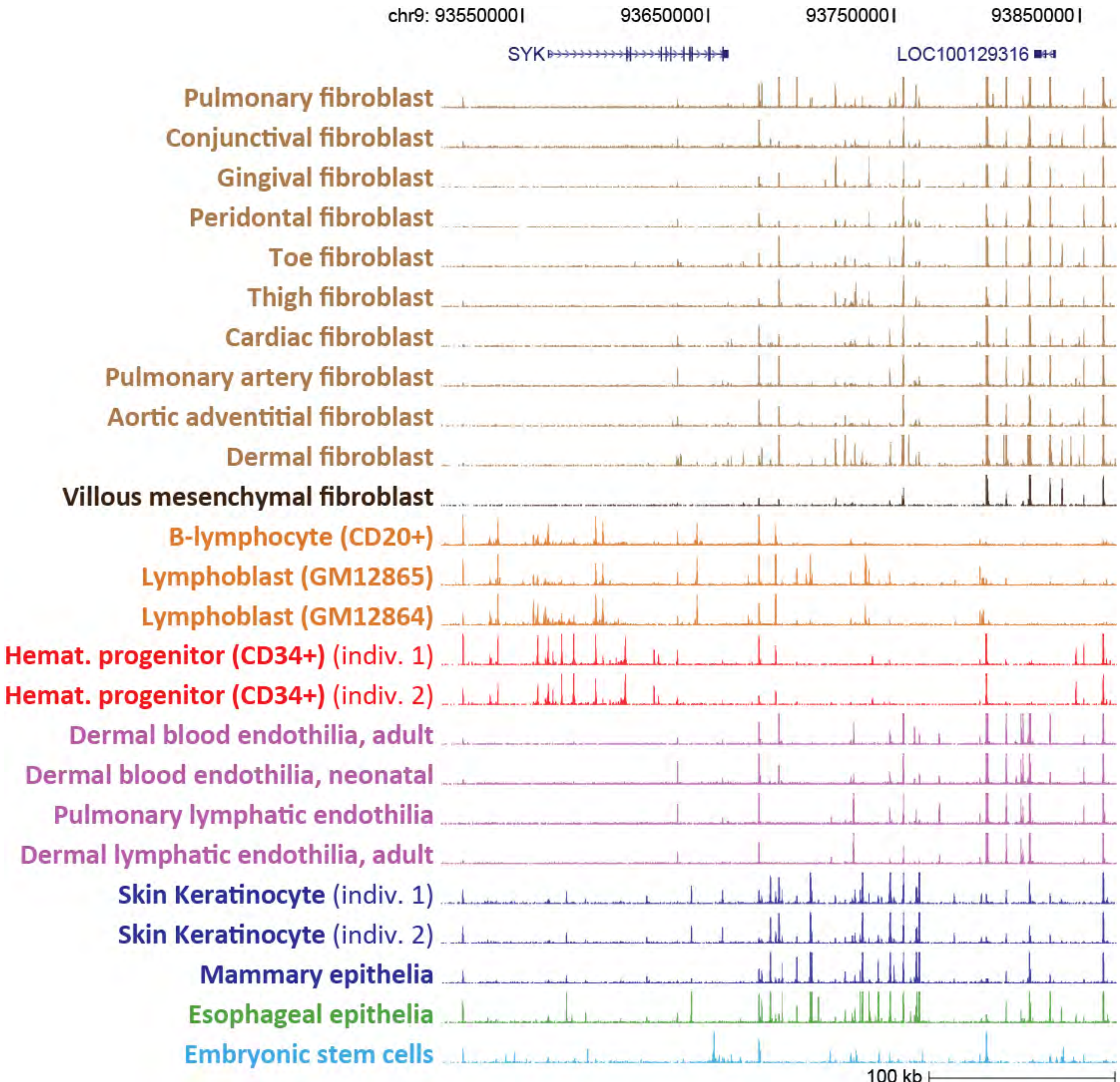


Actuation of regulatory DNA is highly cell-selective

Regulatory regions are highly cell- and lineage-selective



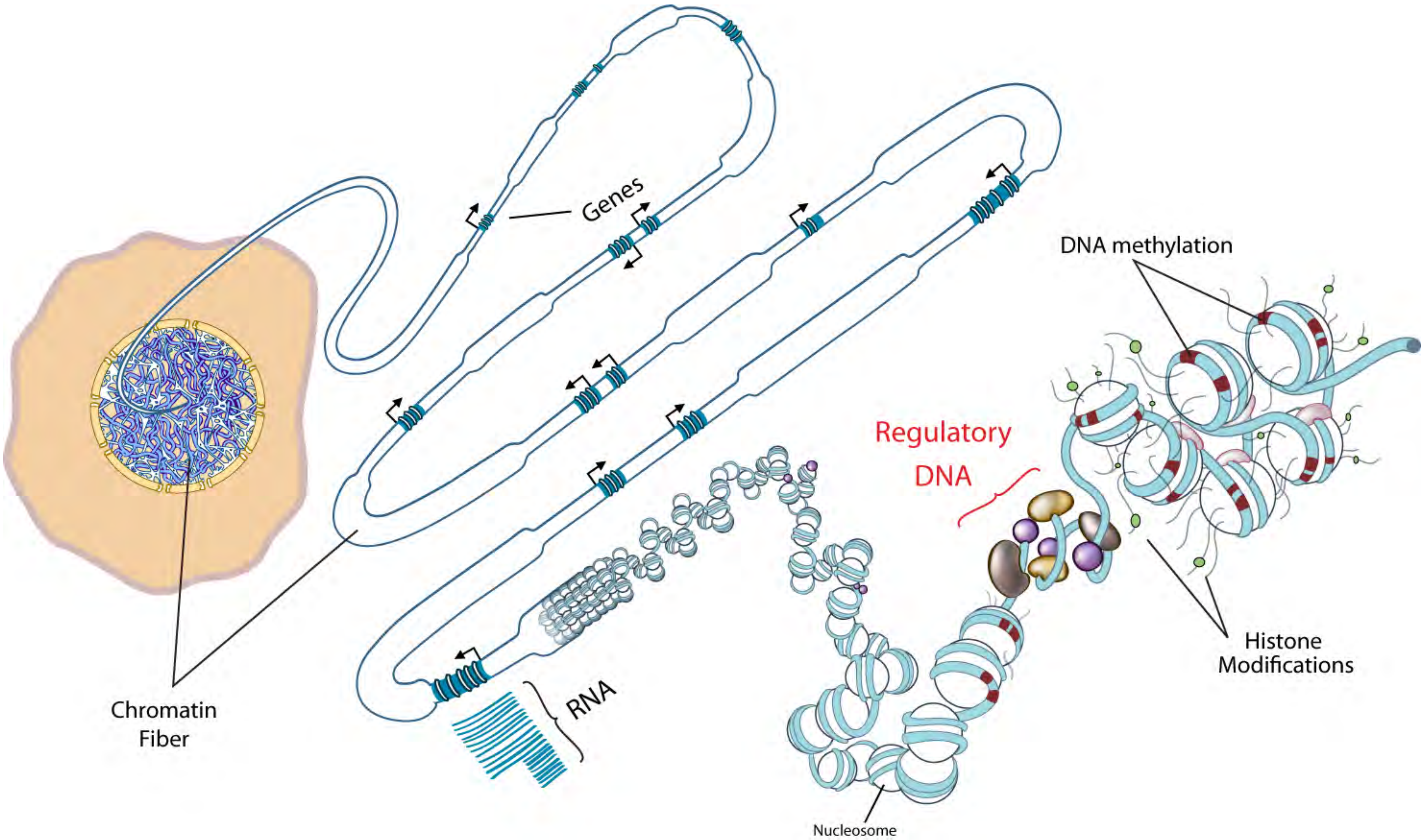
Regulatory regions are highly cell- and lineage-selective



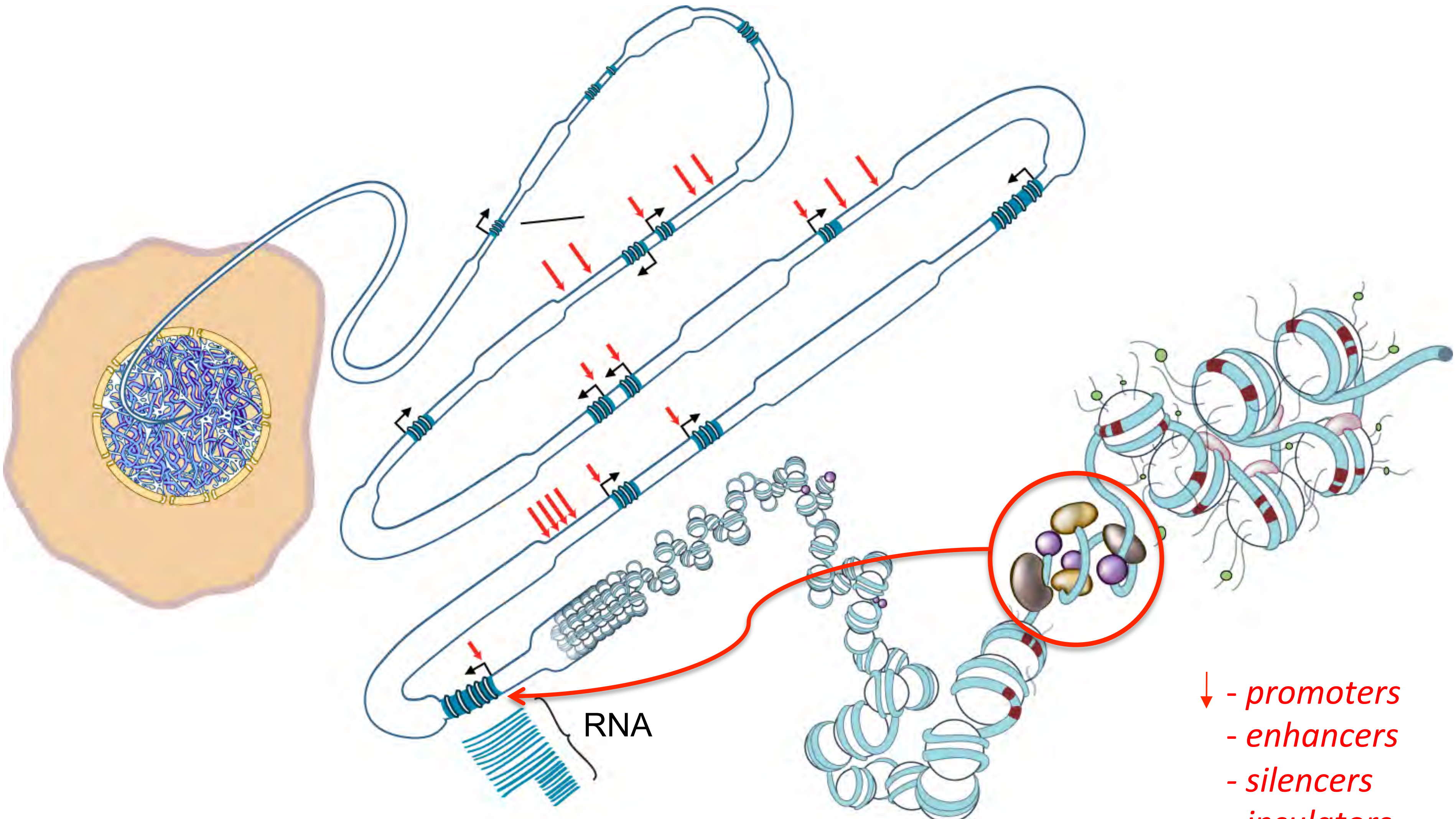
Stergachis et al Cell 2013

Regulatory DNA and genes are highly interconnected

The living genome

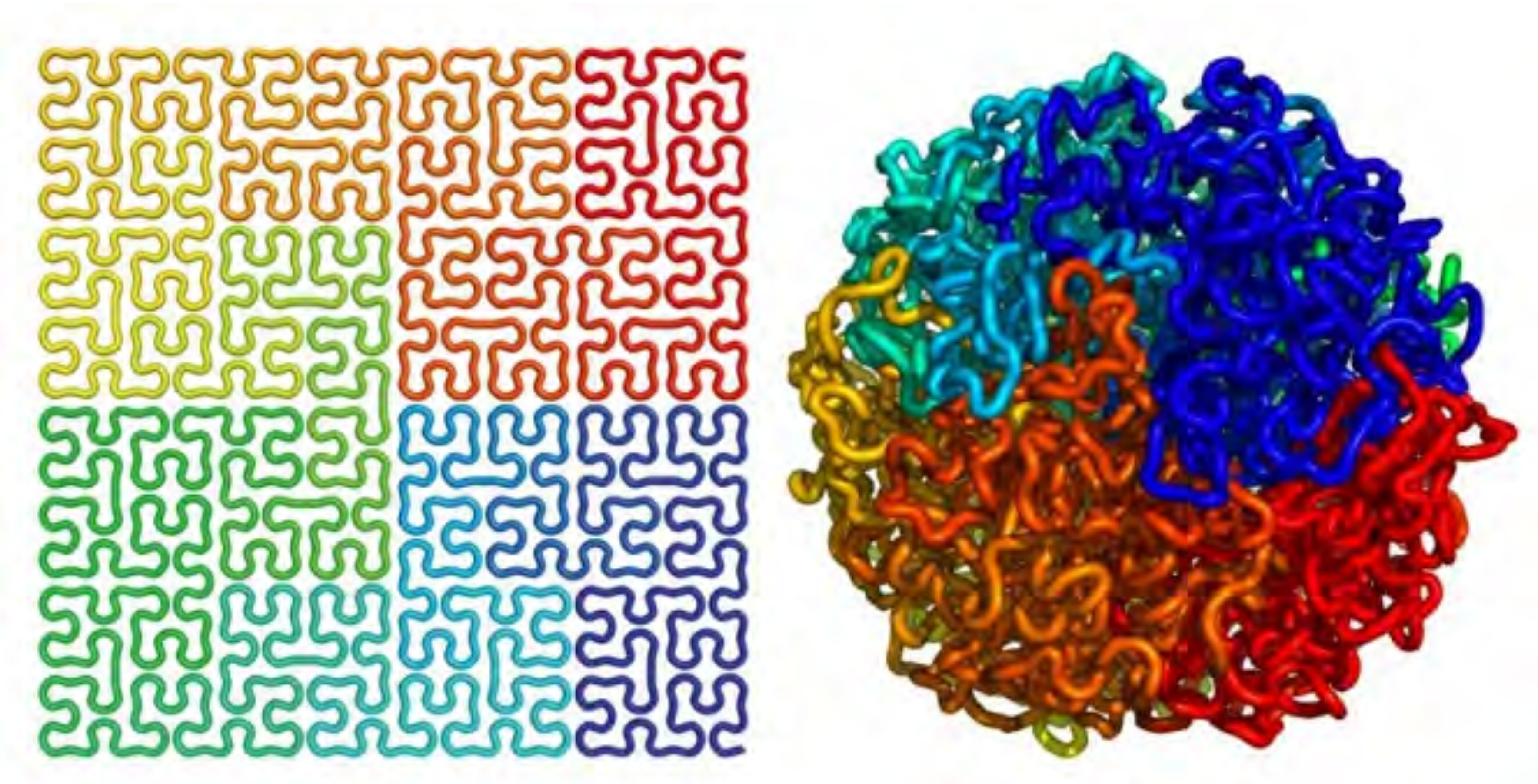


The living genome



↓ - *promoters*
- *enhancers*
- *silencers*
- *insulators*
- *etc.*

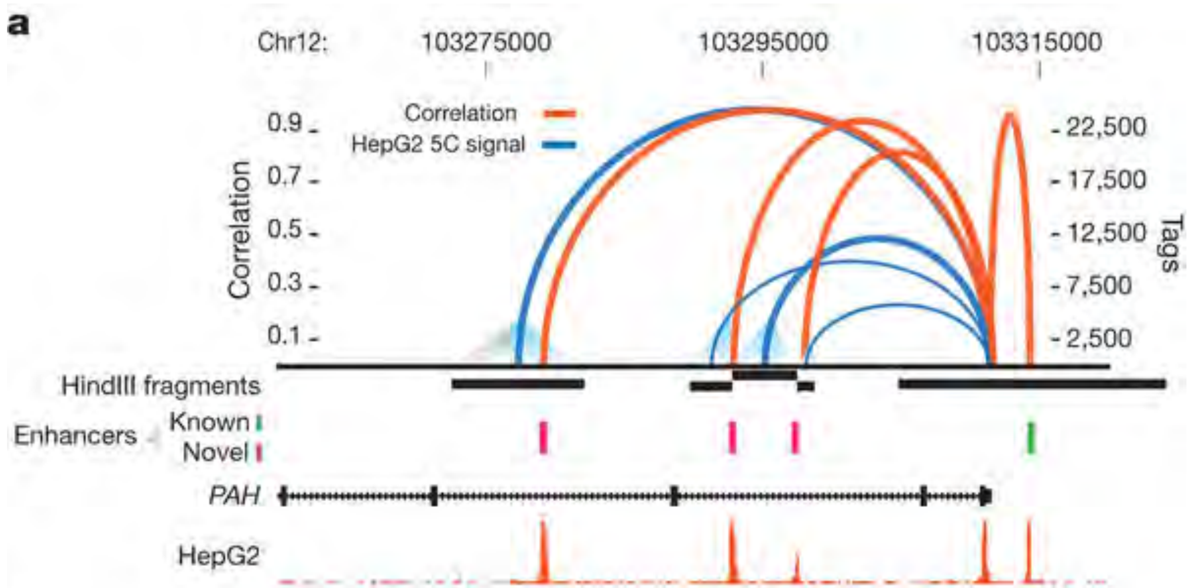
The living genome is fundamentally non-linear

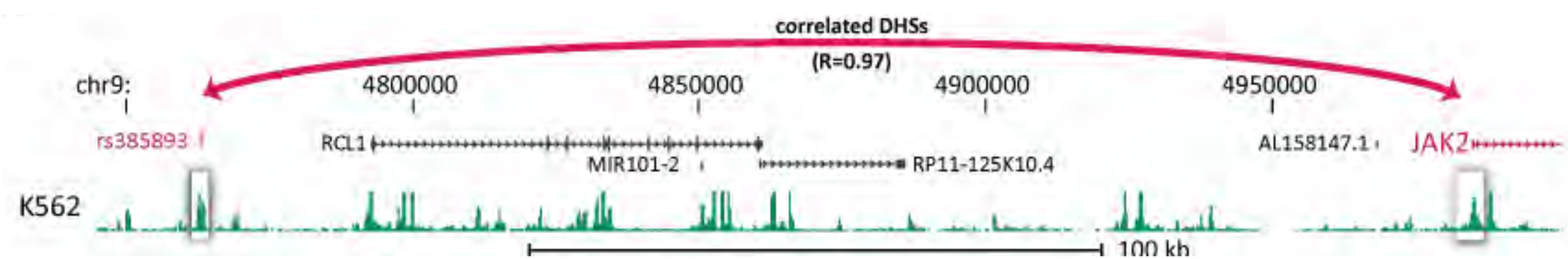


Genomic distance ≠ actual distance

Liberman / Mirny / Dekker Science 2009

Regulatory DNA and genes are densely interconnected *in cis*





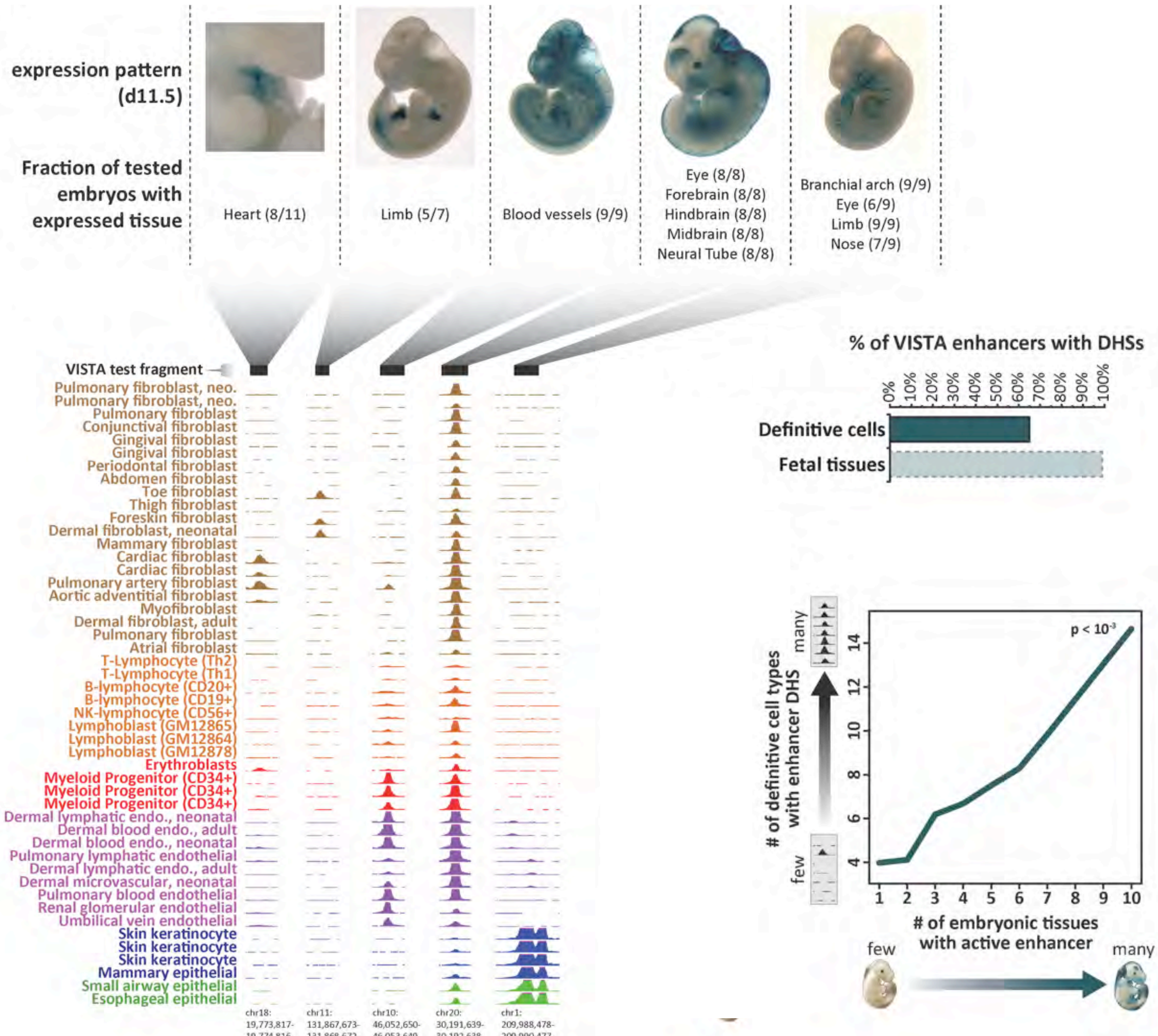
The average enhancer is connected to 1-2 genes

The average gene (promoter) is connected to 15-20 enhancers

Thurman et al. Nature 2012
Maurano et a., Science 2012

Developmentally persistent accessibility at regulatory DNA provides 'memory' of prior cell states

Developmental persistence of enhancer accessibility



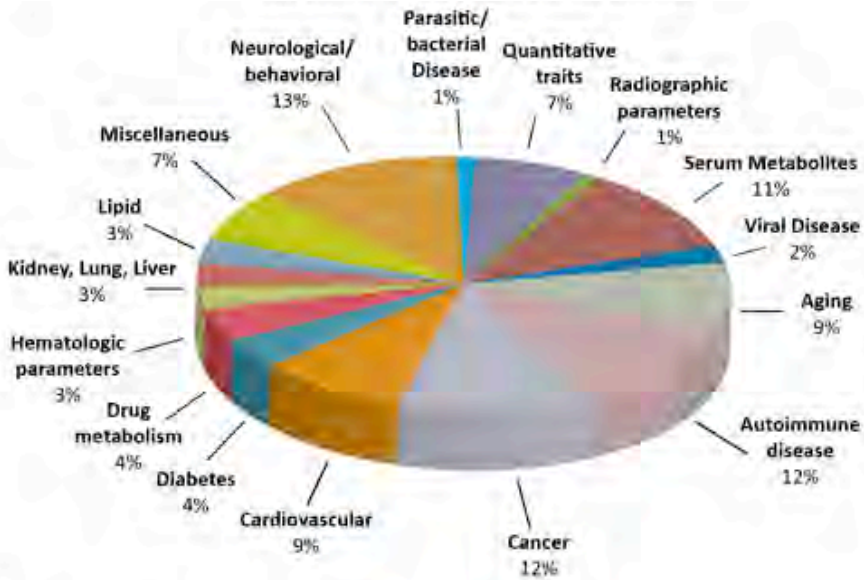
Stergachis et al Cell 2013

Regulatory DNA variation associated with common diseases and traits

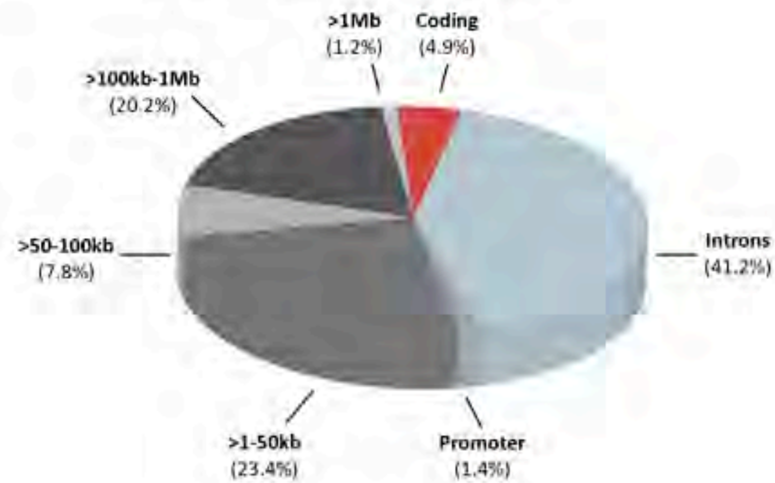
All genetic variation is interpreted in an epigenetic context

Identification of disease- and trait-associated variation by GWAS

GWAS Studies



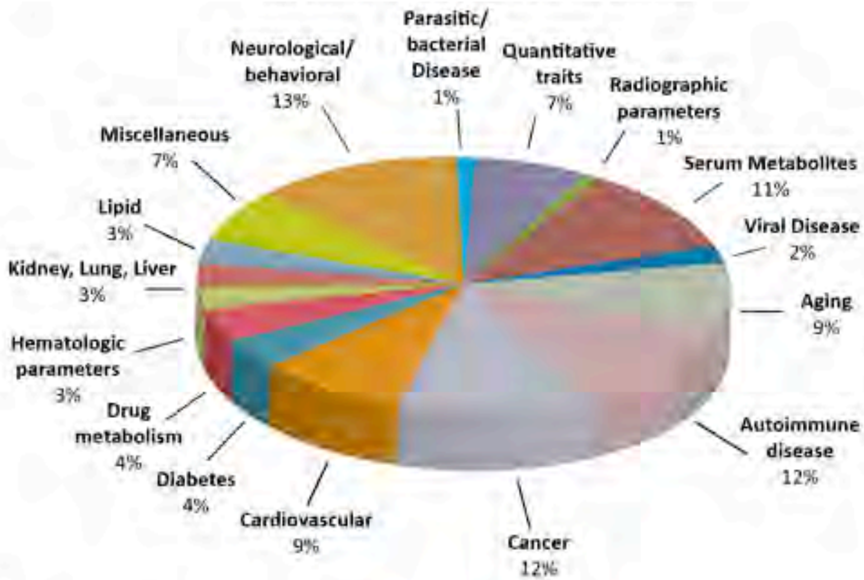
Distribution of GWAS SNPs vs. genes



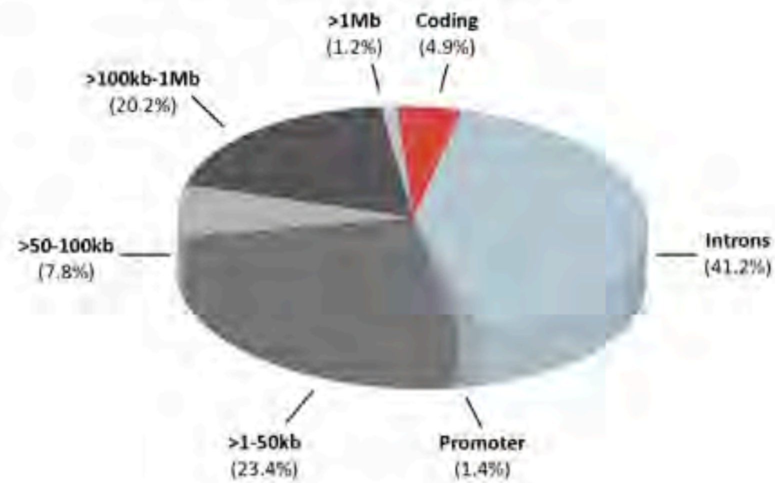
Maurano et al., Science 2012

Identification of disease- and trait-associated variation by GWAS

GWAS Studies



Distribution of GWAS SNPs vs. genes



GWAS disease/trait associated variants

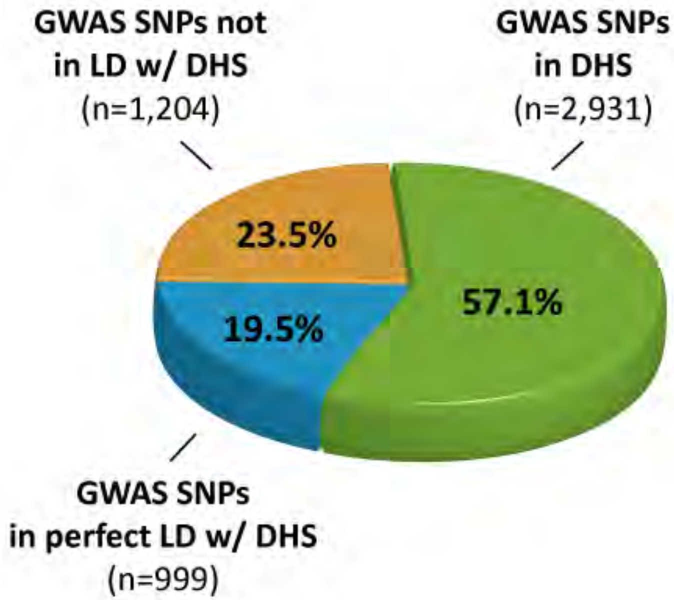
x

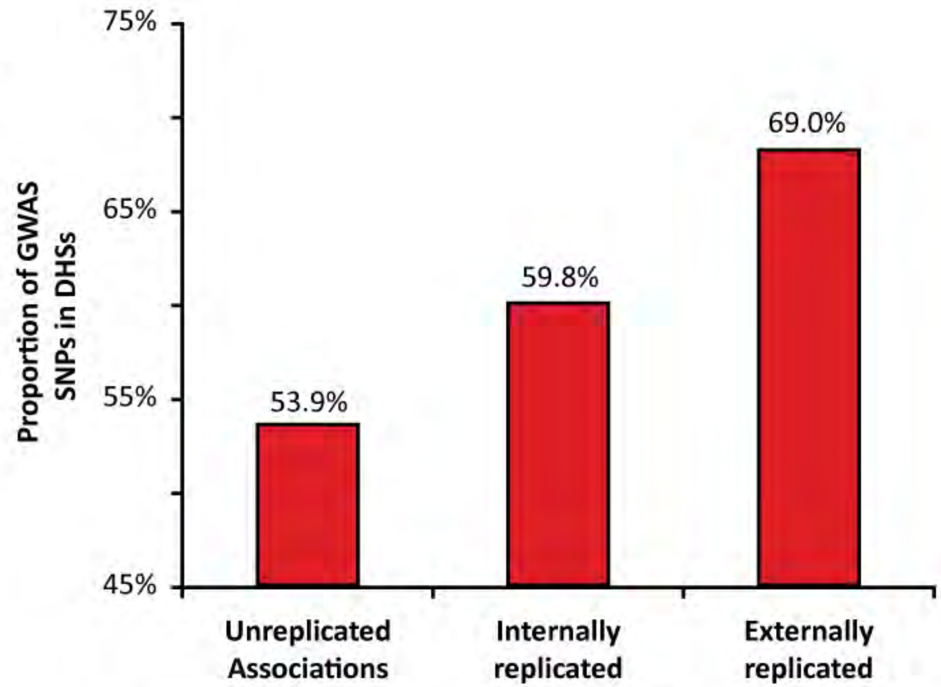
Maps of regulatory DNA in >300 diverse cell and tissue types

Maurano et al., Science 2012

Disease-associated variation is concentrated in regulatory DNA

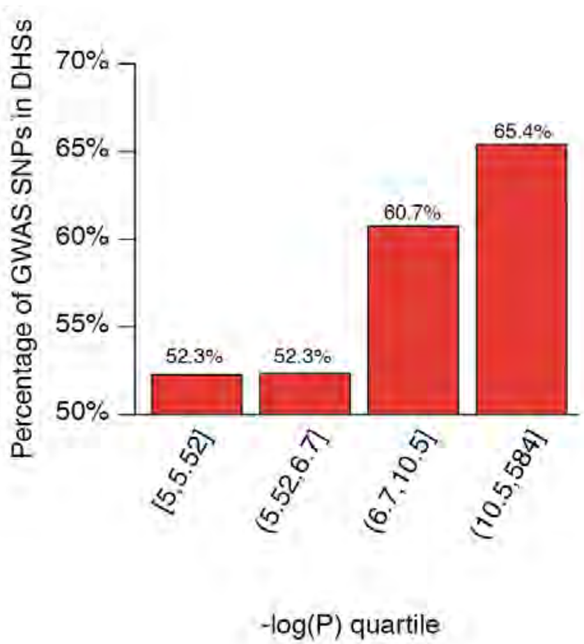
Disease- and trait-associated SNPs are concentrated in regulatory DNA

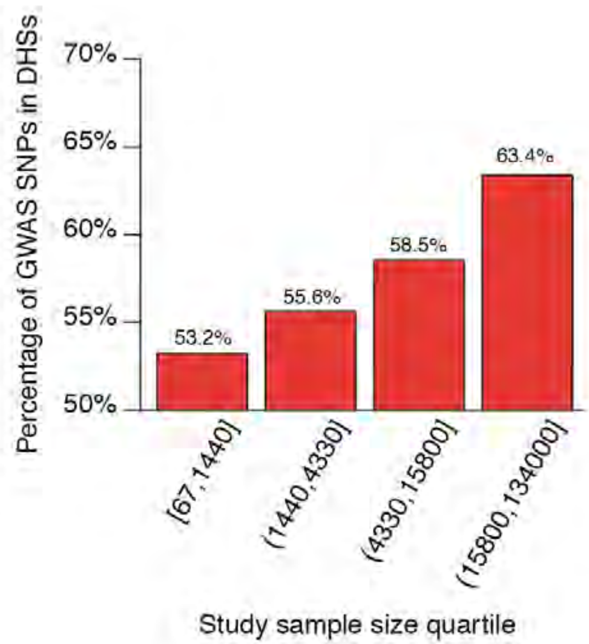




Maurano et al., Science 2012

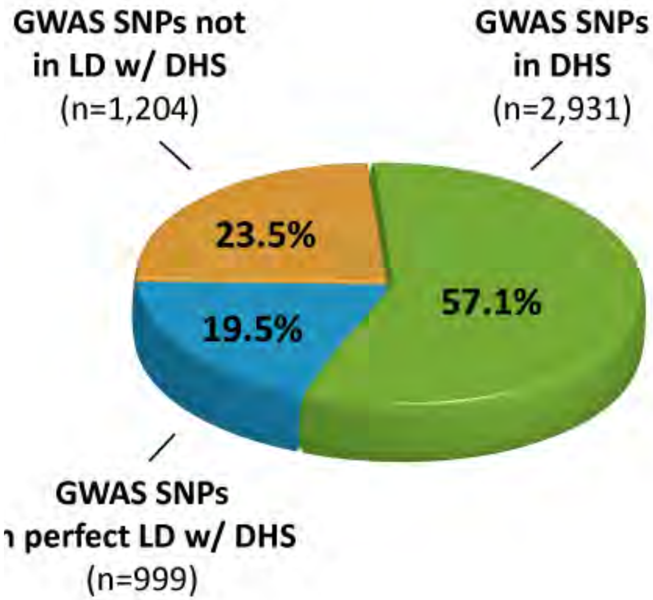
The effect increases monotonically with other measures of higher quality associations

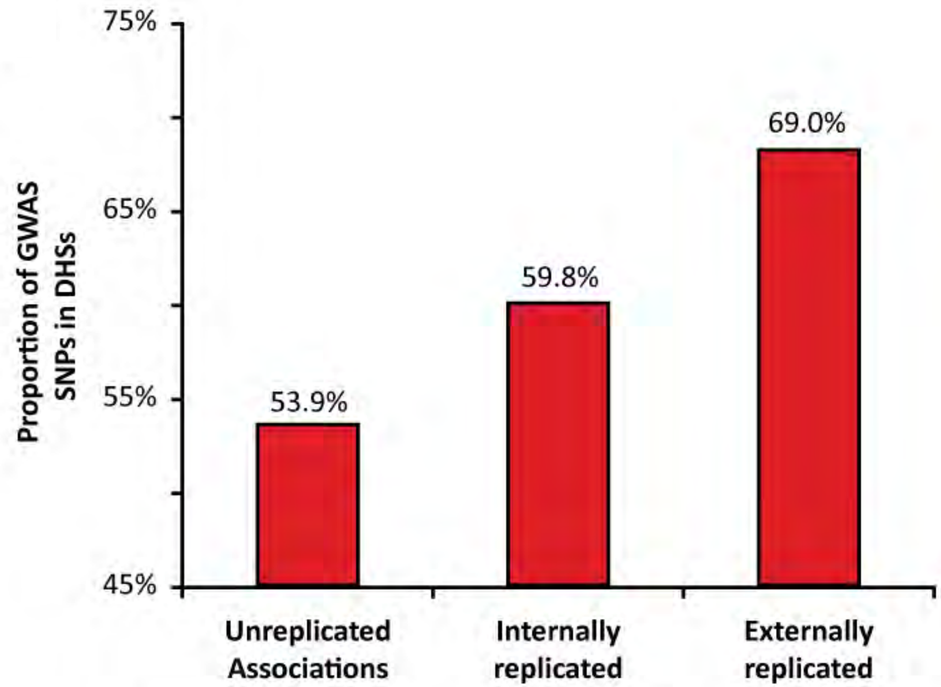




Maurano et al., Science 2012

Disease- and trait-associated SNPs are concentrated in regulatory DNA



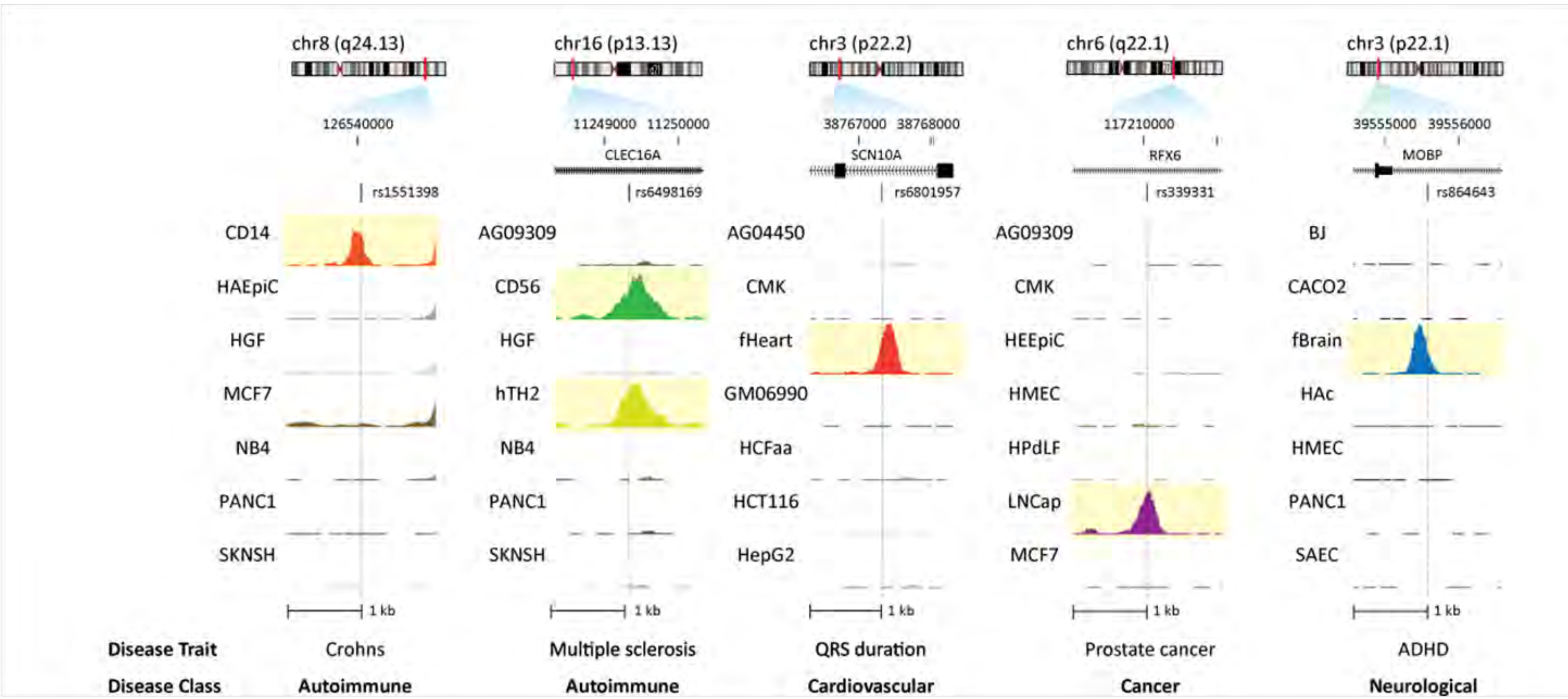


~1.8-fold for all replicated variants in all disorders

>10-fold for specific disese-cell type pairings

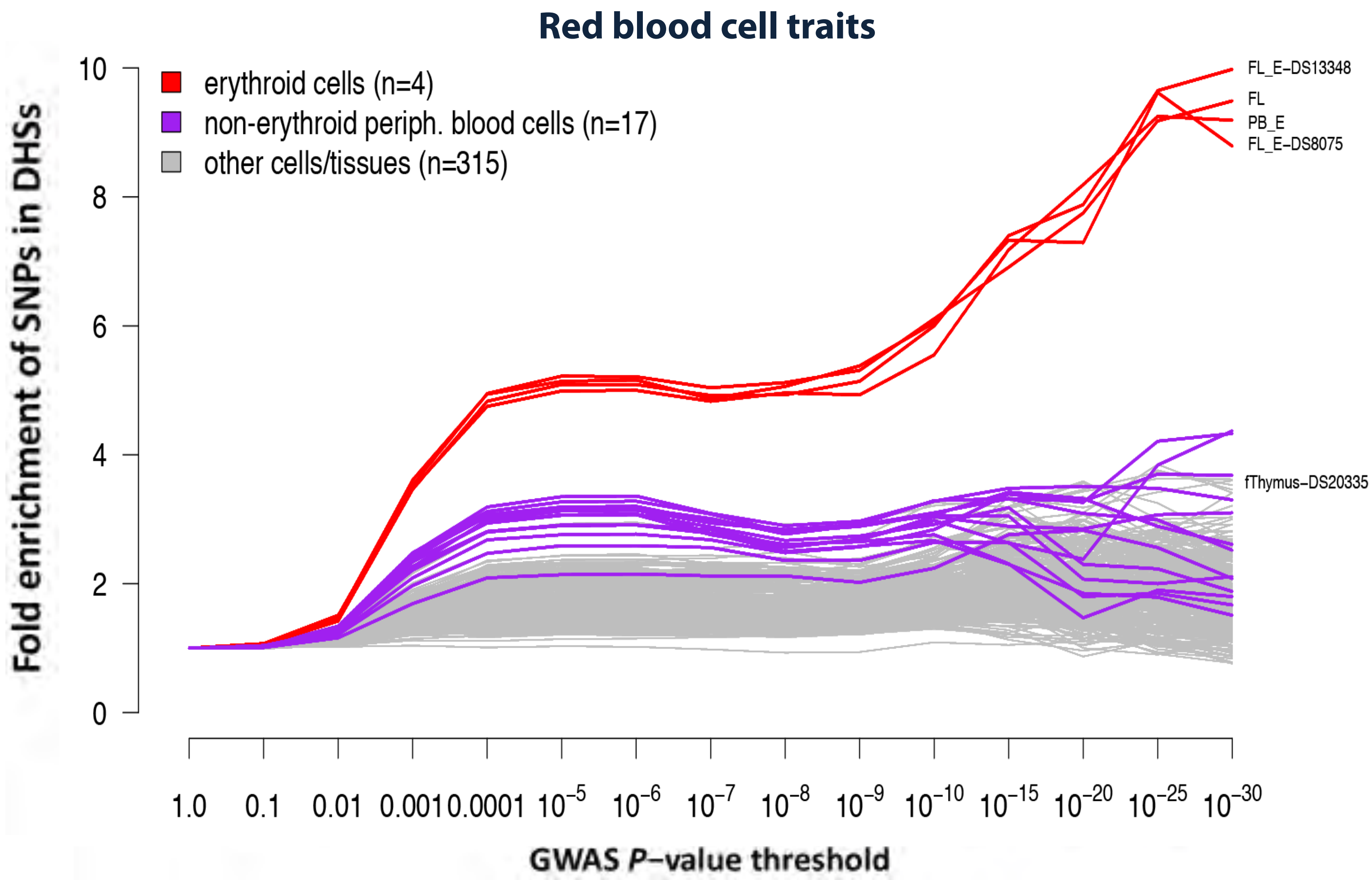
GWAS variants selectively localize in regulatory DNA of pathologically relevant cell types

Disease-associated variation clusters in pathogenic or target cell types



Maurano et al., Science 2012

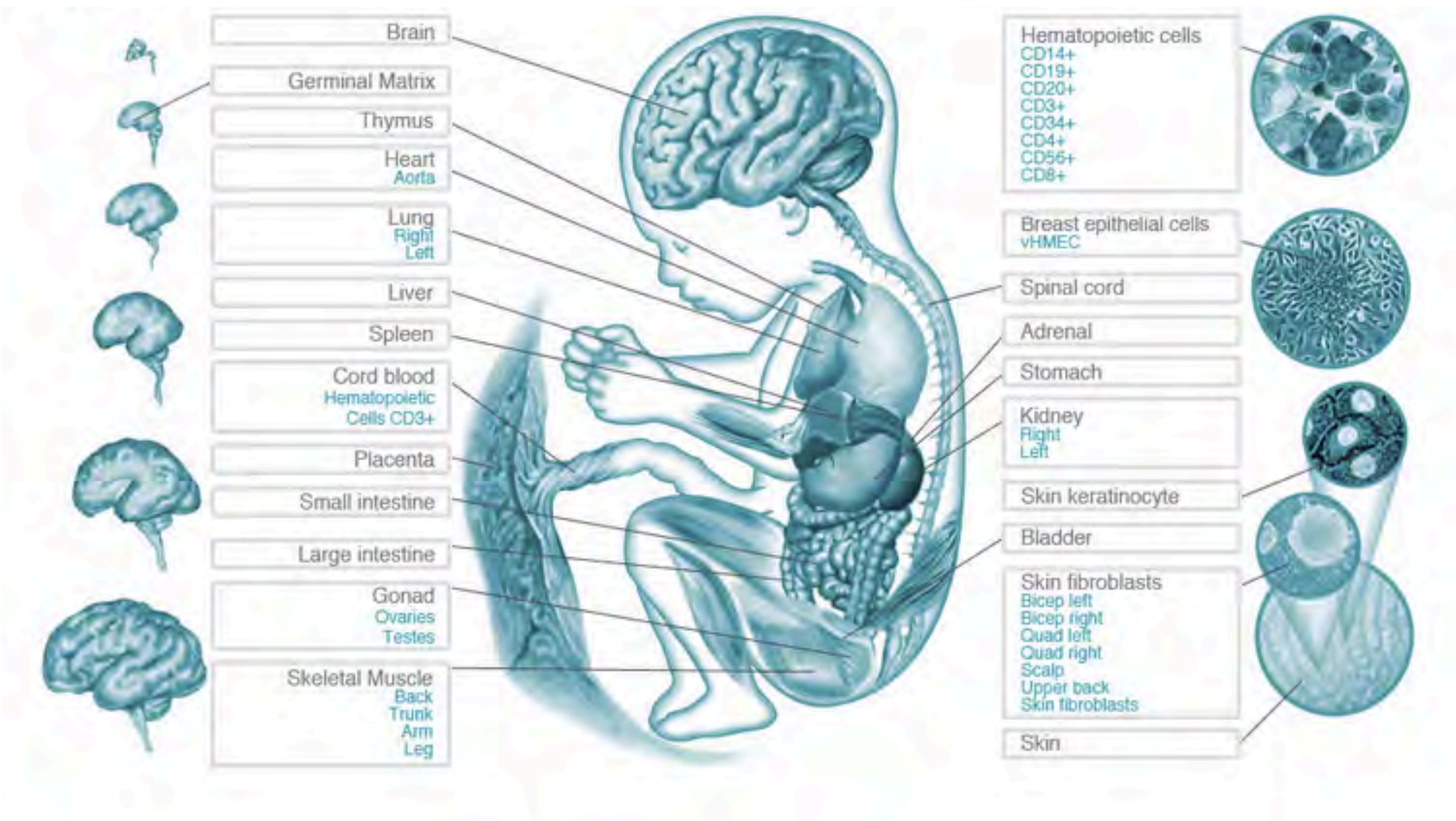
Cell-selective enrichment of trait-associated variants



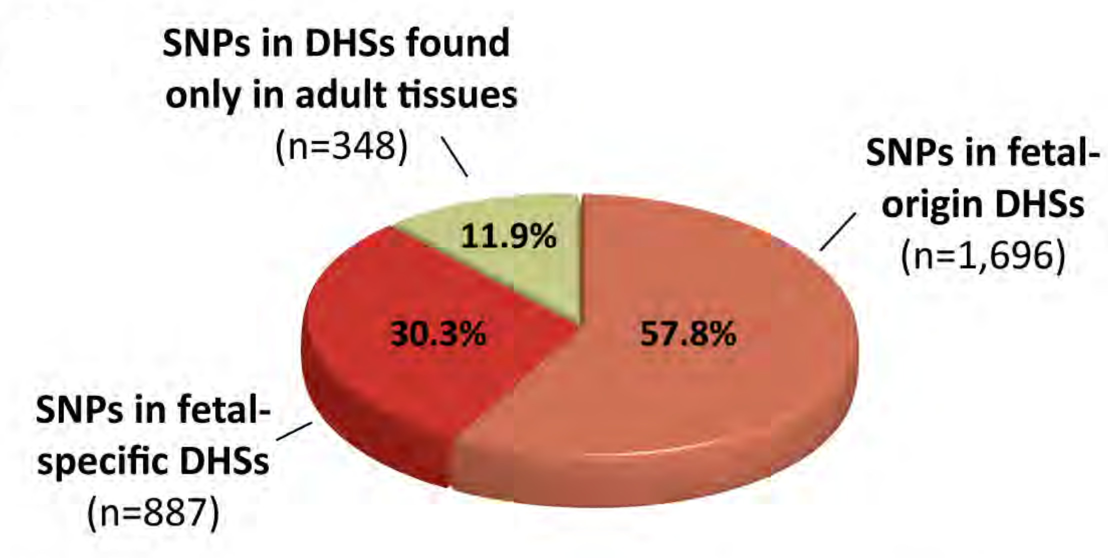
Variants associated with diseases and traits with developmental contributions preferentially localize in fetal regulatory DNA

Surveying the normal epigenomic landscape

Developing cells and tissues

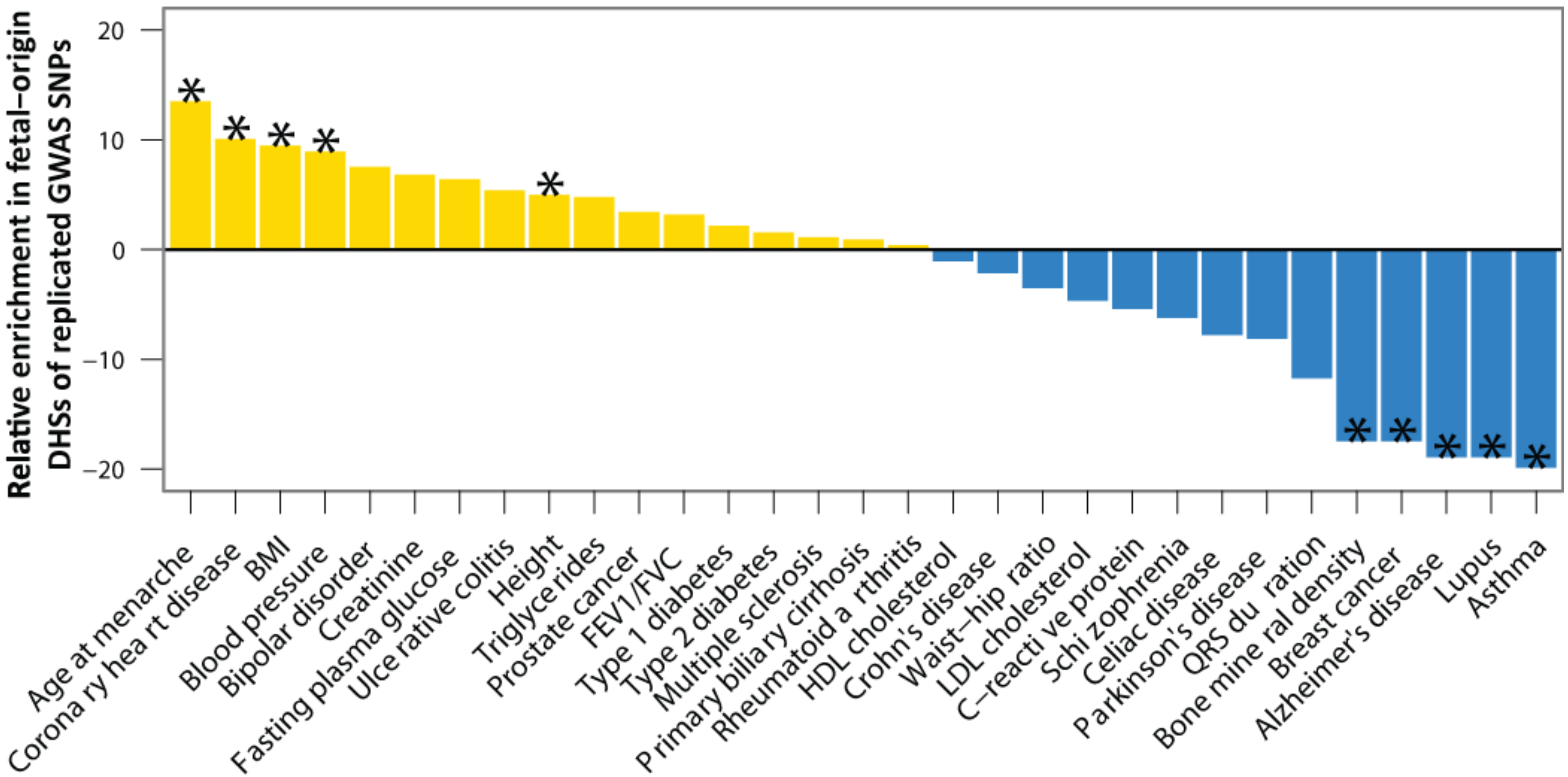


Most variants lie in regulatory DNA of fetal origin



Maurano et al., Science 2012

Fetal regulatory variants are enriched in traits & diseases with known links to intrauterine exposures

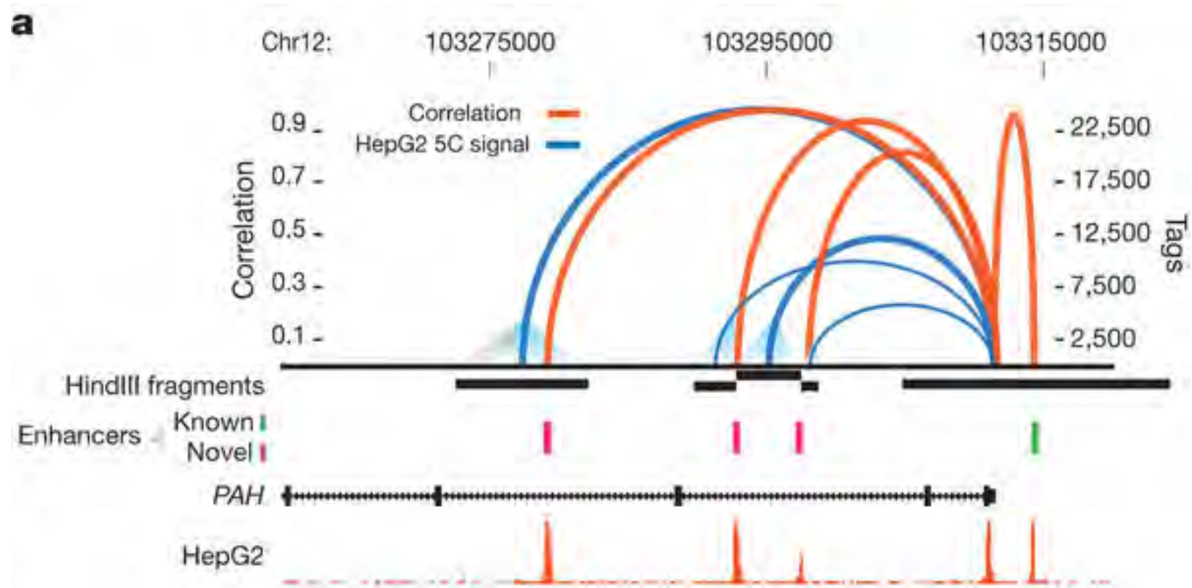


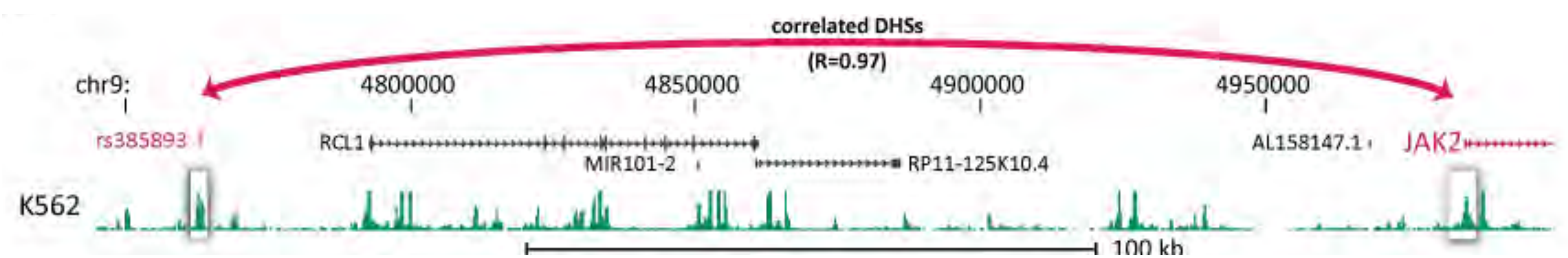
Maurano et al., Science 2012

Correcting genetic associations for epigenetic circuitry

Regulatory DNA harboring disease-associated variants mainly controls distant genes

The epigenome and genes are densely interconnected *in cis*

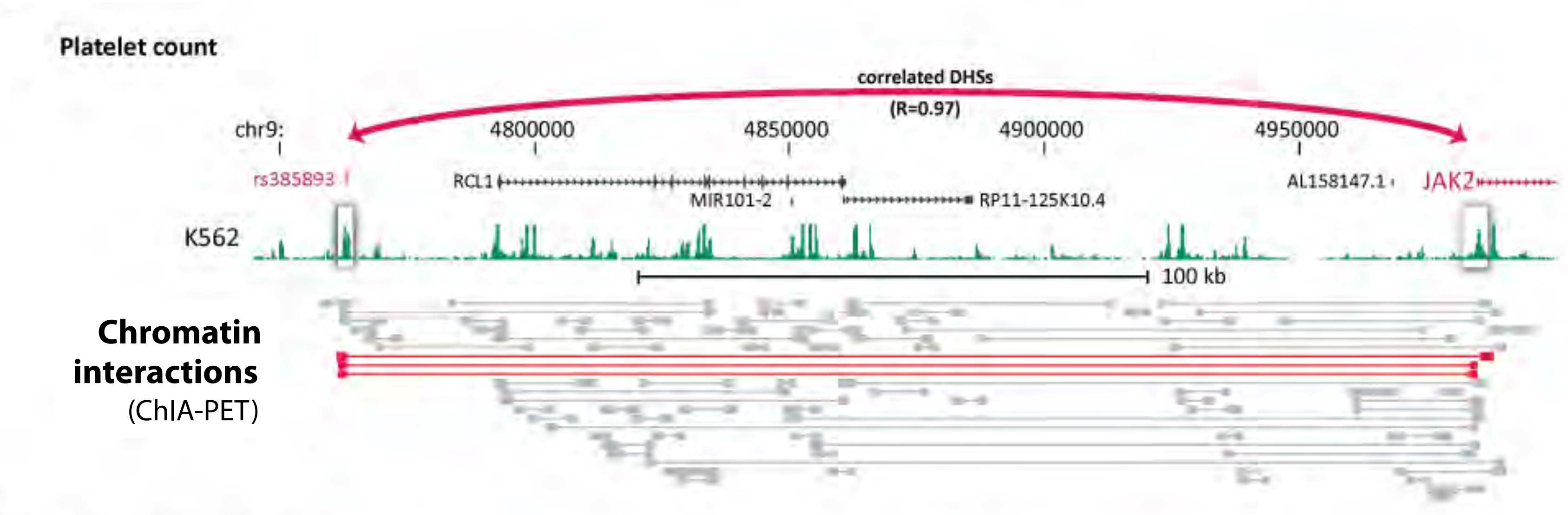




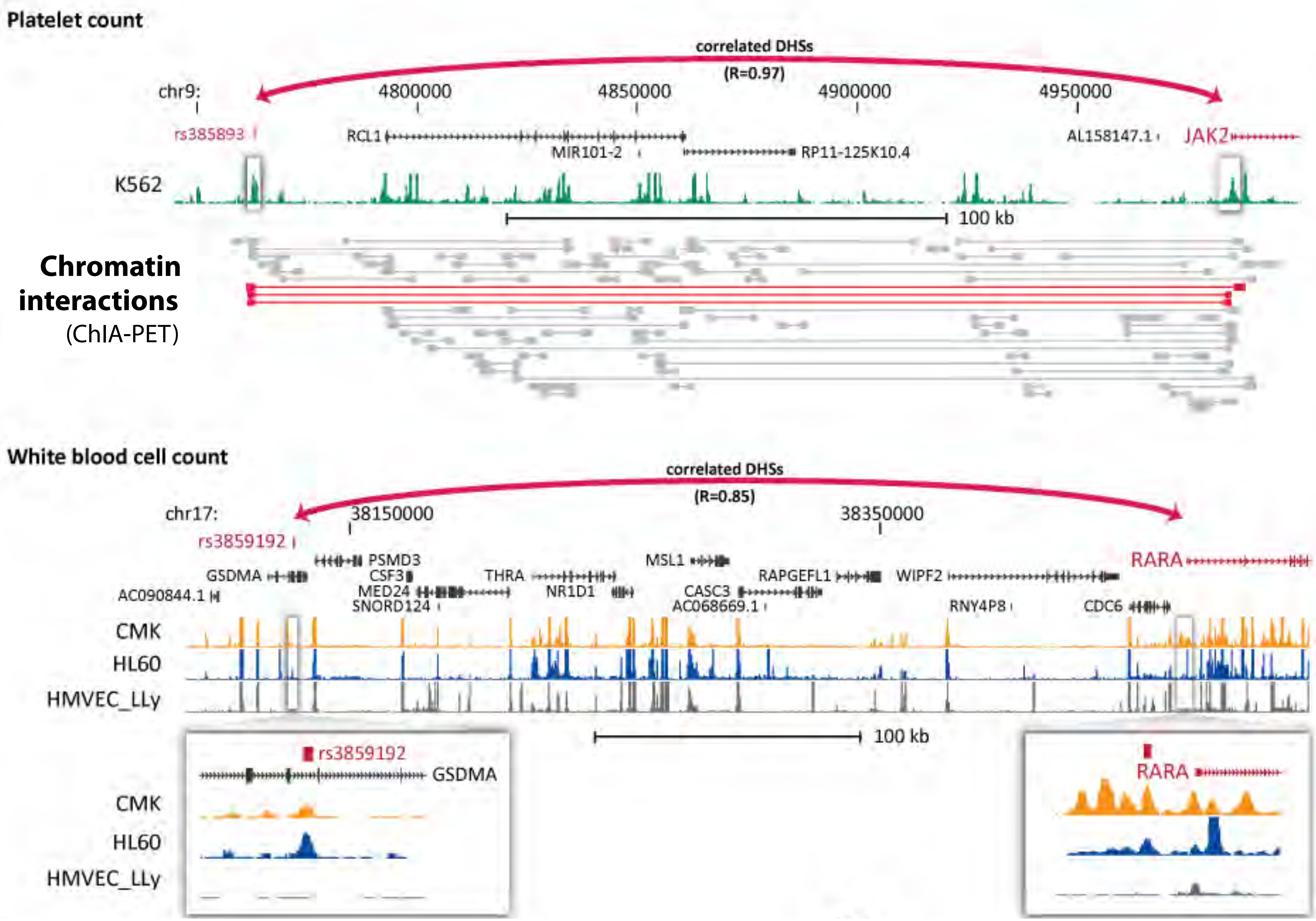
The average enhancer is connected to 1-2 genes
The average gene (promoter) is connected to 15-20 enhancers

Thurman et al. Nature 2012
Maurano et a., Science 2012

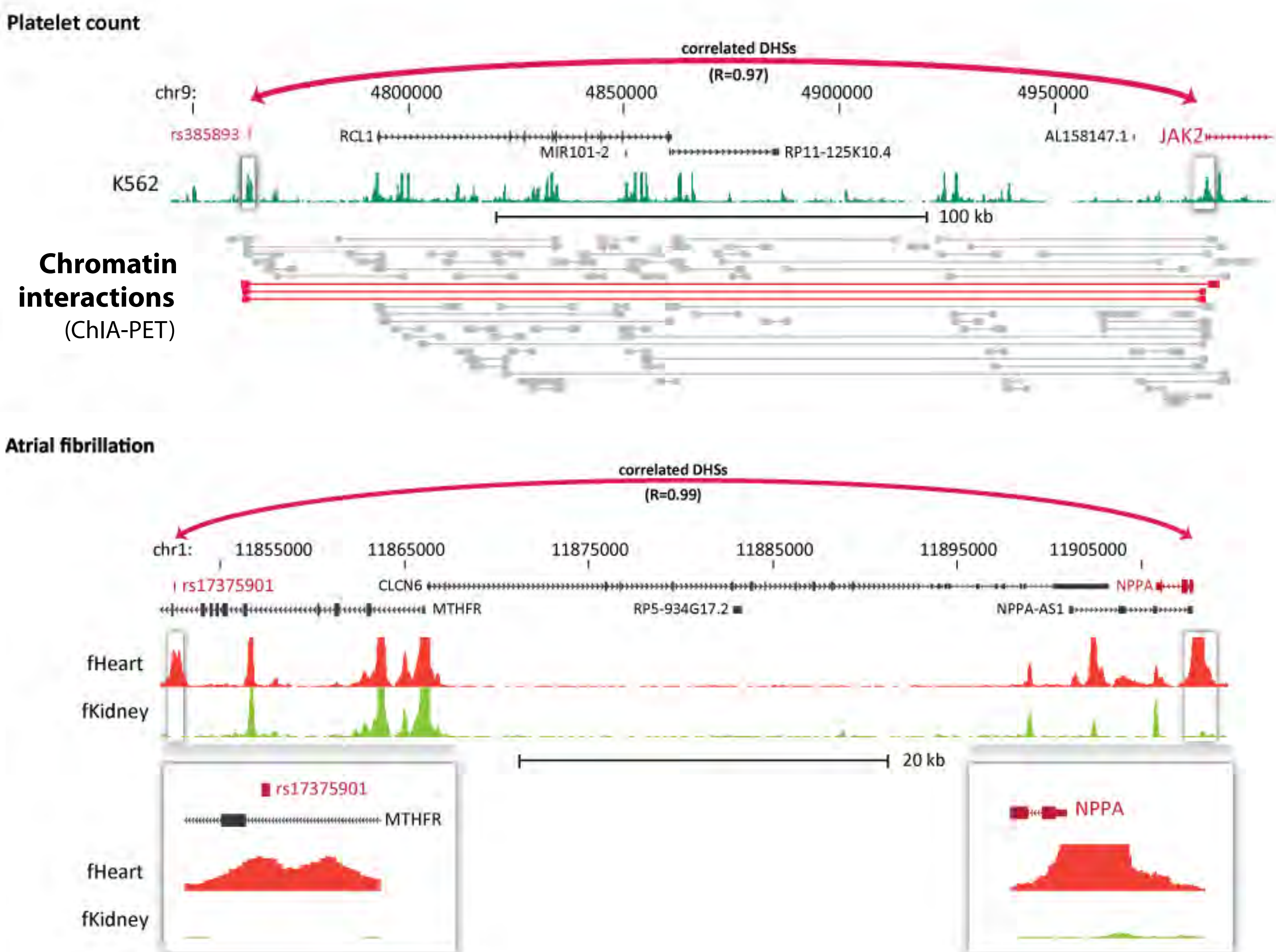
Regulatory GWAS variants linked to distant genes with causative potential



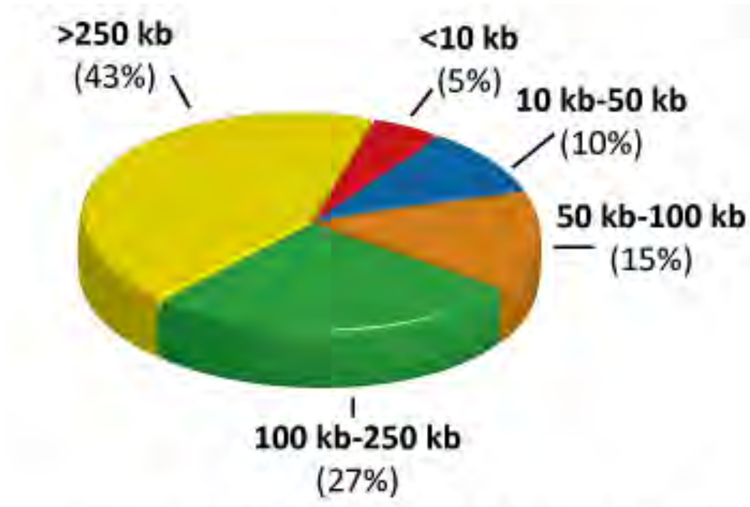
Regulatory GWAS variants linked to distant genes with causative potential



Regulatory GWAS variants linked to distant genes with causative potential



Regulatory GWAS variants linked to distant genes with pathogenic potential

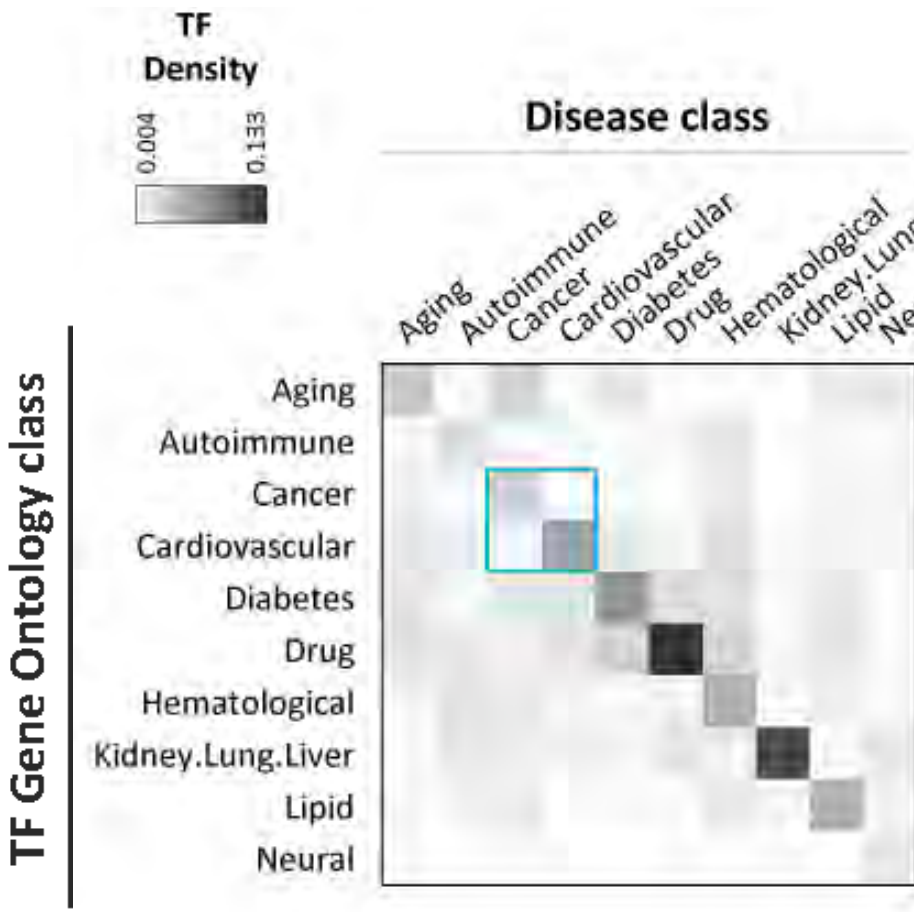


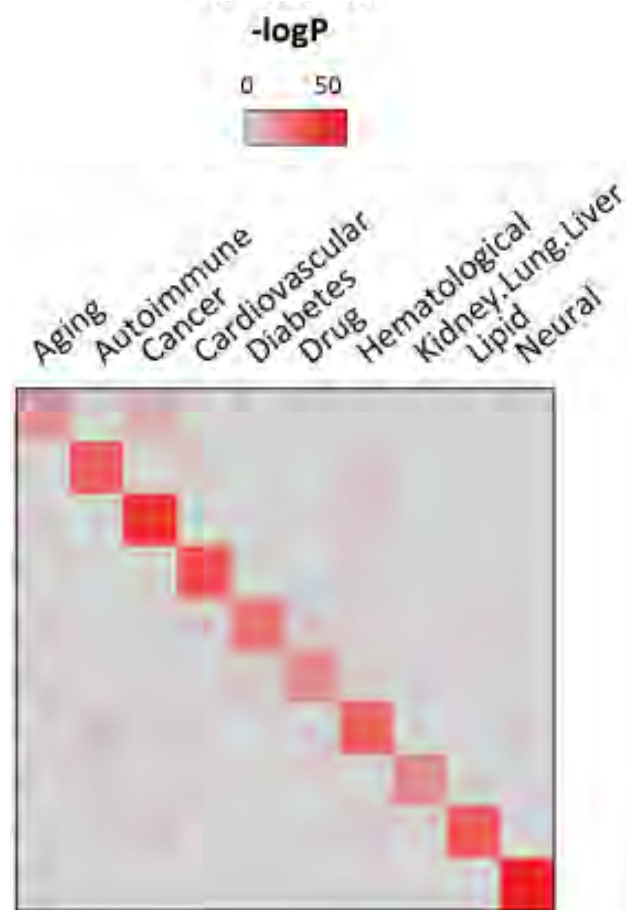
Disease or trait	*r*	Target gene	Function	Distance (kb)
Amyotrophic lateral sclerosis	1	SYNGAP1*	Axon formation; component of NMDA complex	411
Crohn's disease	1	TRIB1*	NF-κB regulation	95
Time to first primary tooth	0.99	PRDM1*	Craniofacial development	452
C-reactive protein	0.99	NLRP3	Response to bacterial pathogens	20
Multiple sclerosis	0.98	AHI1*	White matter abnormalities	149
QRS duration	0.96	SCN10A*	Sodium channel involved in cardiac conduction	181
Breast cancer	0.96	TACC2*	Tumor suppressor	411
Schizophrenia/brain imaging	0.95	KIF1A*	Neuron-specific kinesin involved in axonal transport	428
Brain structure	0.94	CXCR6*	Chemokine receptor involved in glial migration	357
Rheumatoid arthritis	0.94	CTSB*	Cysteine proteinase linked to articular erosion	359
Ovarian cancer	0.93	HSPG2*	Ovarian tumor suppressor	268
Multiple sclerosis	0.93	ZP1*	Known autoantigen	153
ADHD	0.93	PDLIM5*	Neuronal calcium signaling	328
Breast cancer	0.88	MAP3K1*	Response to growth factors	158
Amyotrophic lateral sclerosis	0.88	CNTN4	Neuronal cell adhesion	306
Schizophrenia	0.81	FXR1*	Cognitive function	120
Type 1 diabetes	0.75	ACAD10*	Mitochondrial oxidation of fatty acids	343
Lupus	0.74	STAT4	Mediates IL-12 immune response and T_H1 differentiation	113

Maurano et al., Science 2012

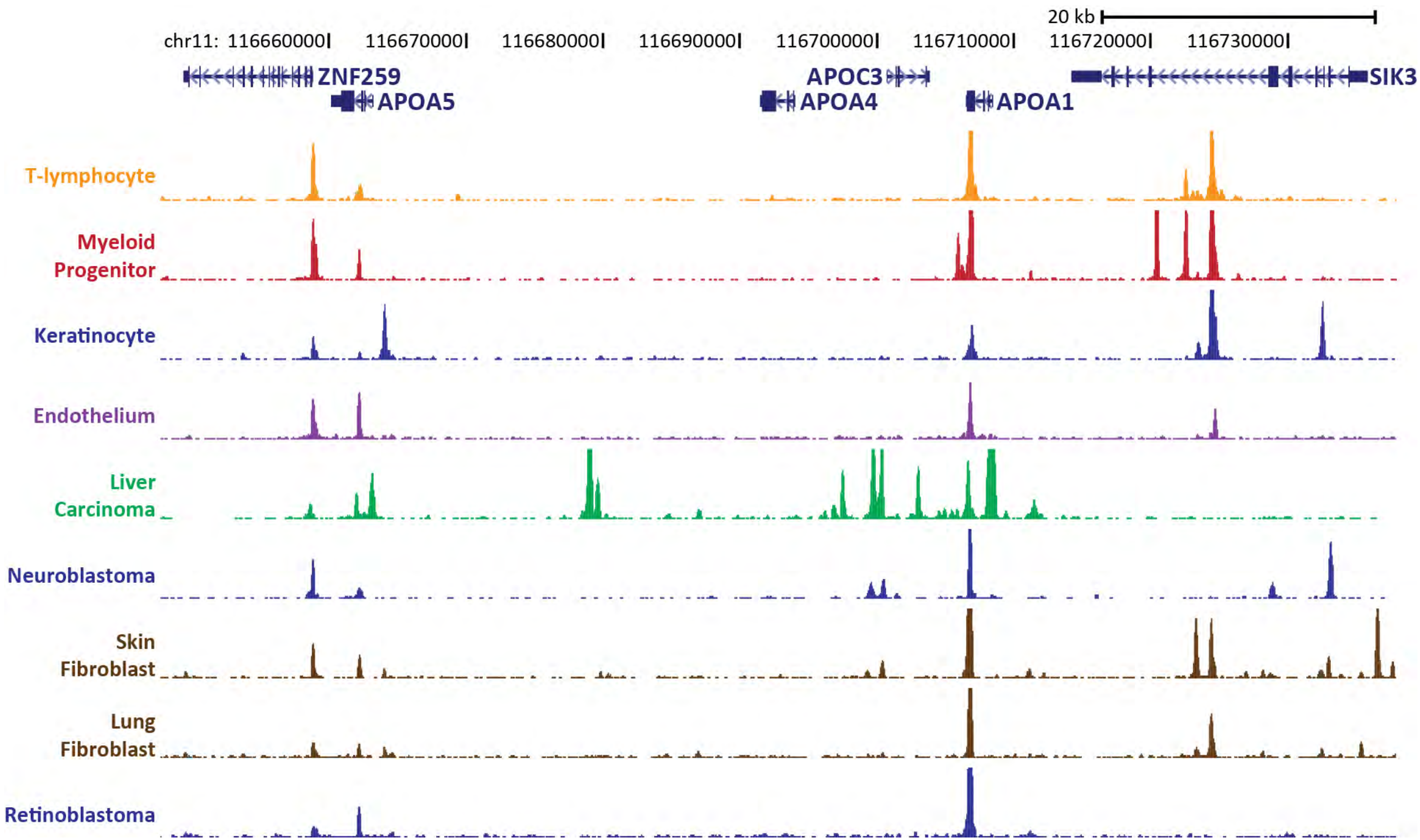
Disease-associated variants selectively localize to relevant transcription factor recognition sites

Within regulatory DNA, disease-associated variants systematically localize within relevant TF recognition sites

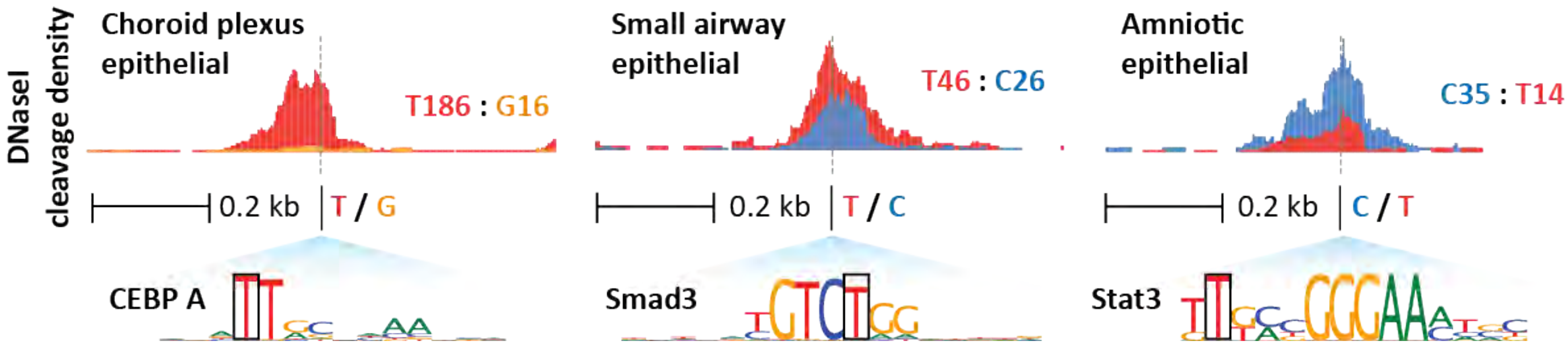




DNaseI hypersensitive sites mark regulatory DNA



Disease/trait variants specify allelic chromatin states



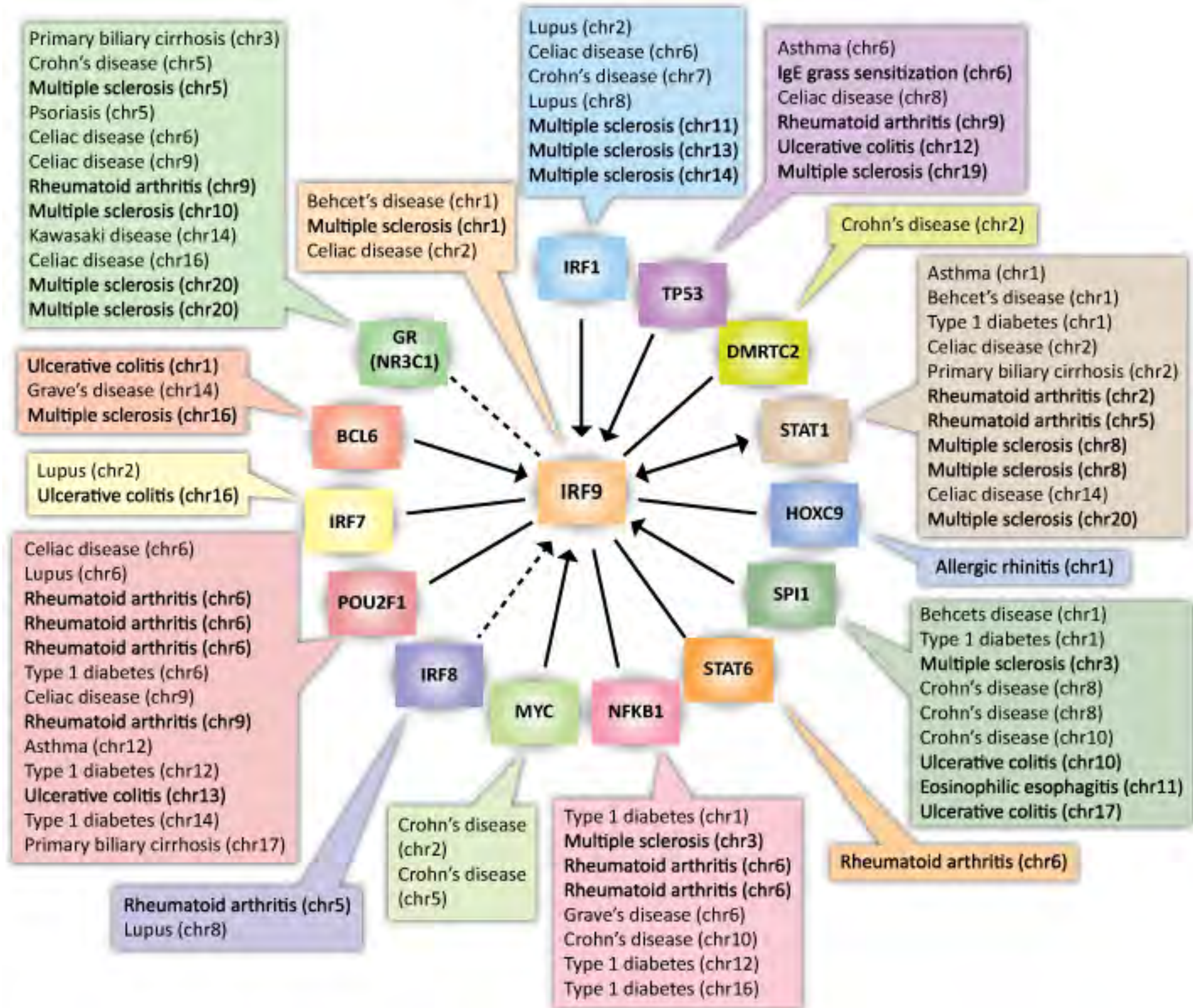
Overall, 20.5% of GWAS SNPs exhibit significant allelic imbalance

For those with high sequencing depth (>200x), 38.7%

Maurano et al., Science 2012

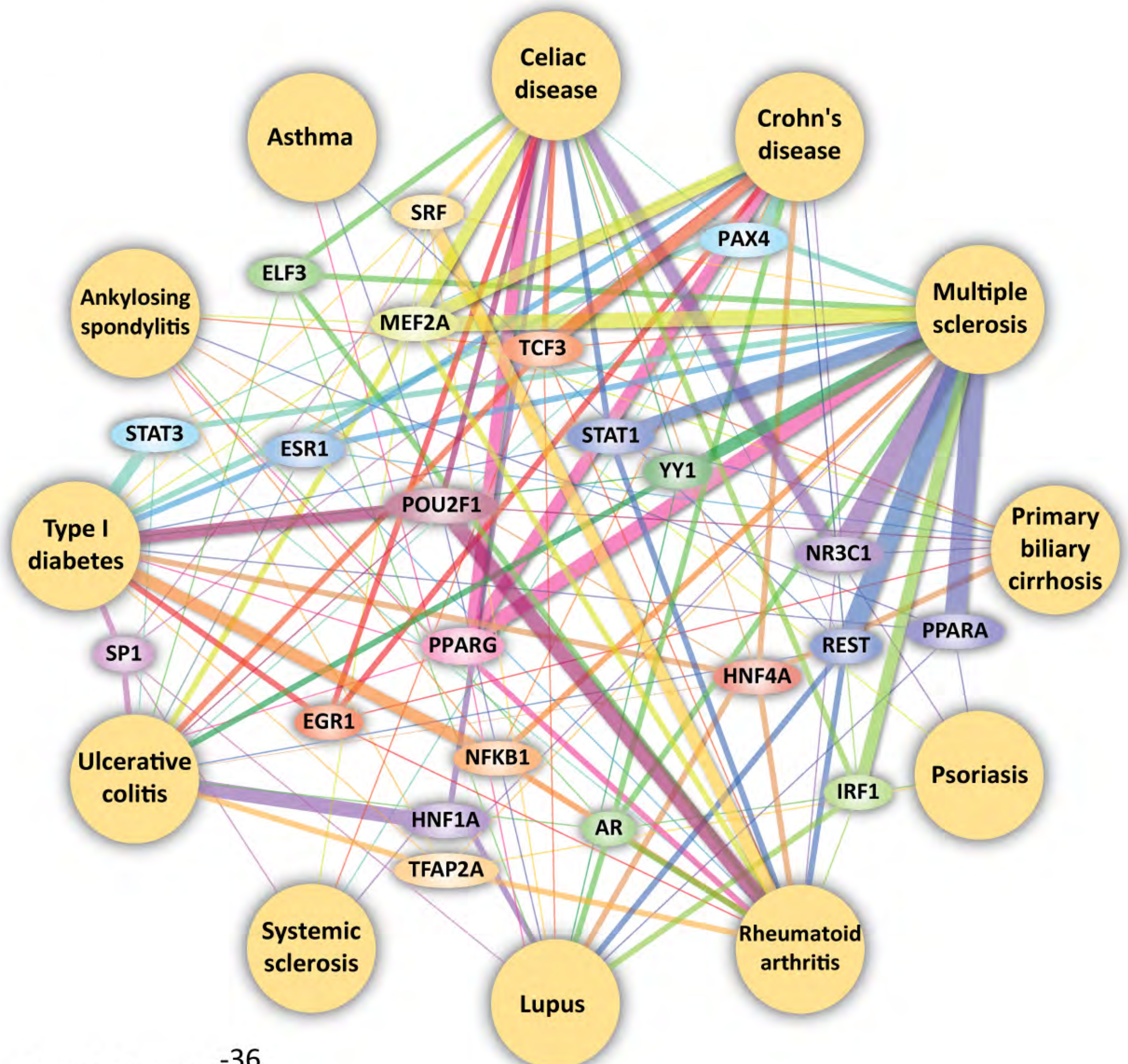
Disease-associated variants cluster in regulatory pathways and form regulatory networks

~25% of inflammatory disease-associated variants in regulatory DNA perturb the JAK/STAT pathway



P < 1.6 x 10^{-13}

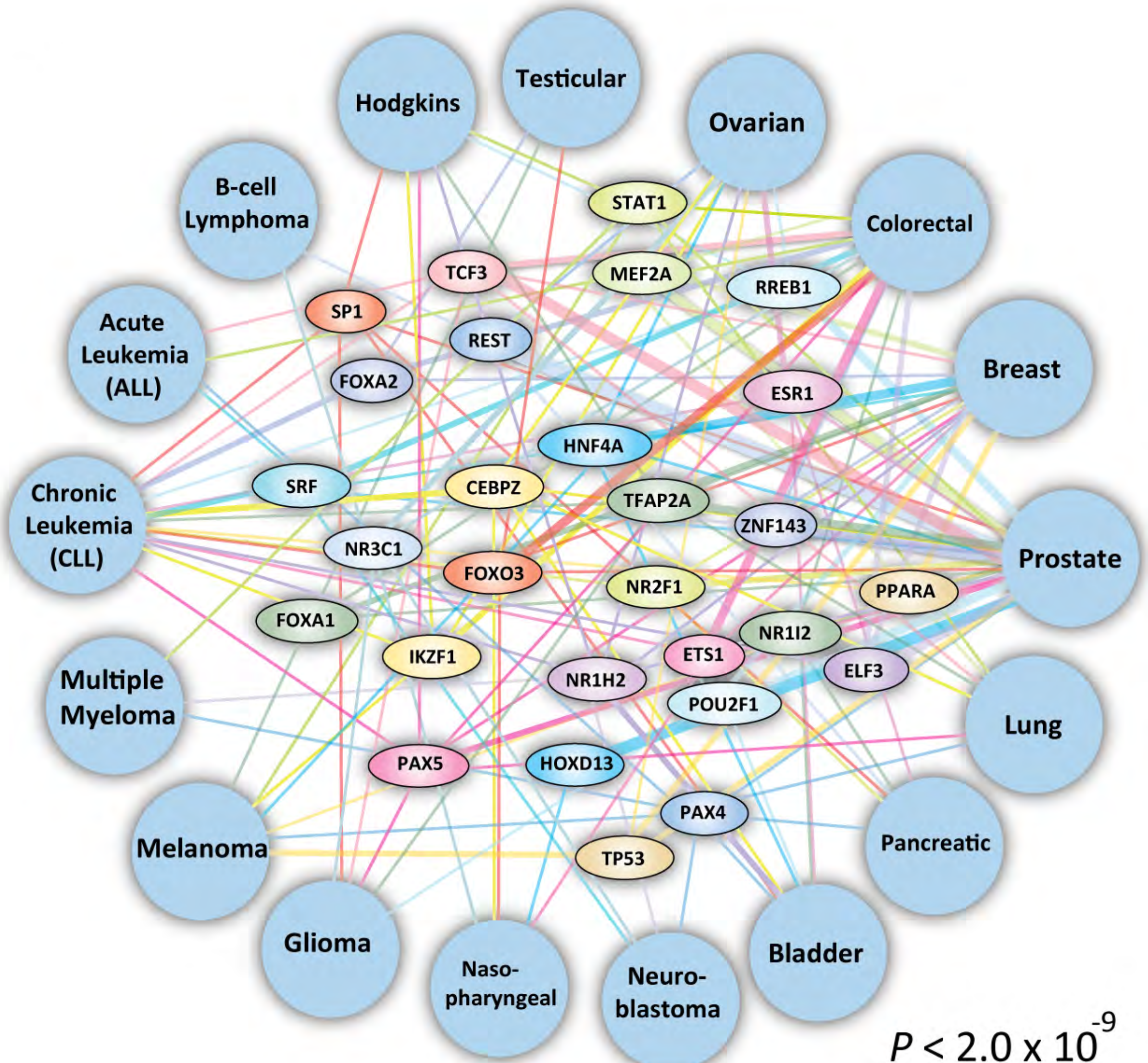
A common regulatory network for autoimmune disease



$P < 1.3 \times 10^{-36}$

Maurano et al., Science 2012

A common regulatory network underlies diverse malignancies

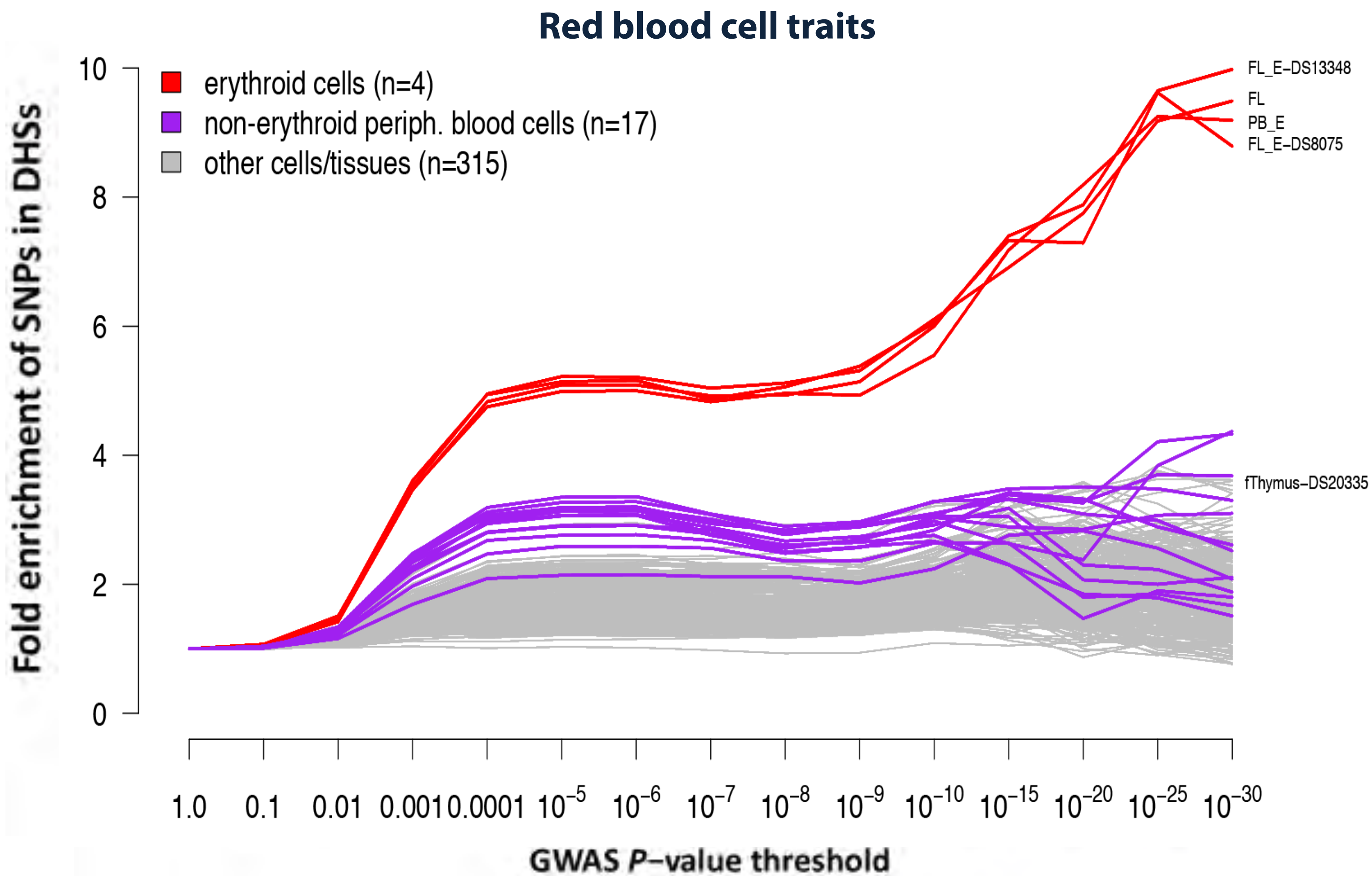


$P < 2.0 \times 10^{-9}$

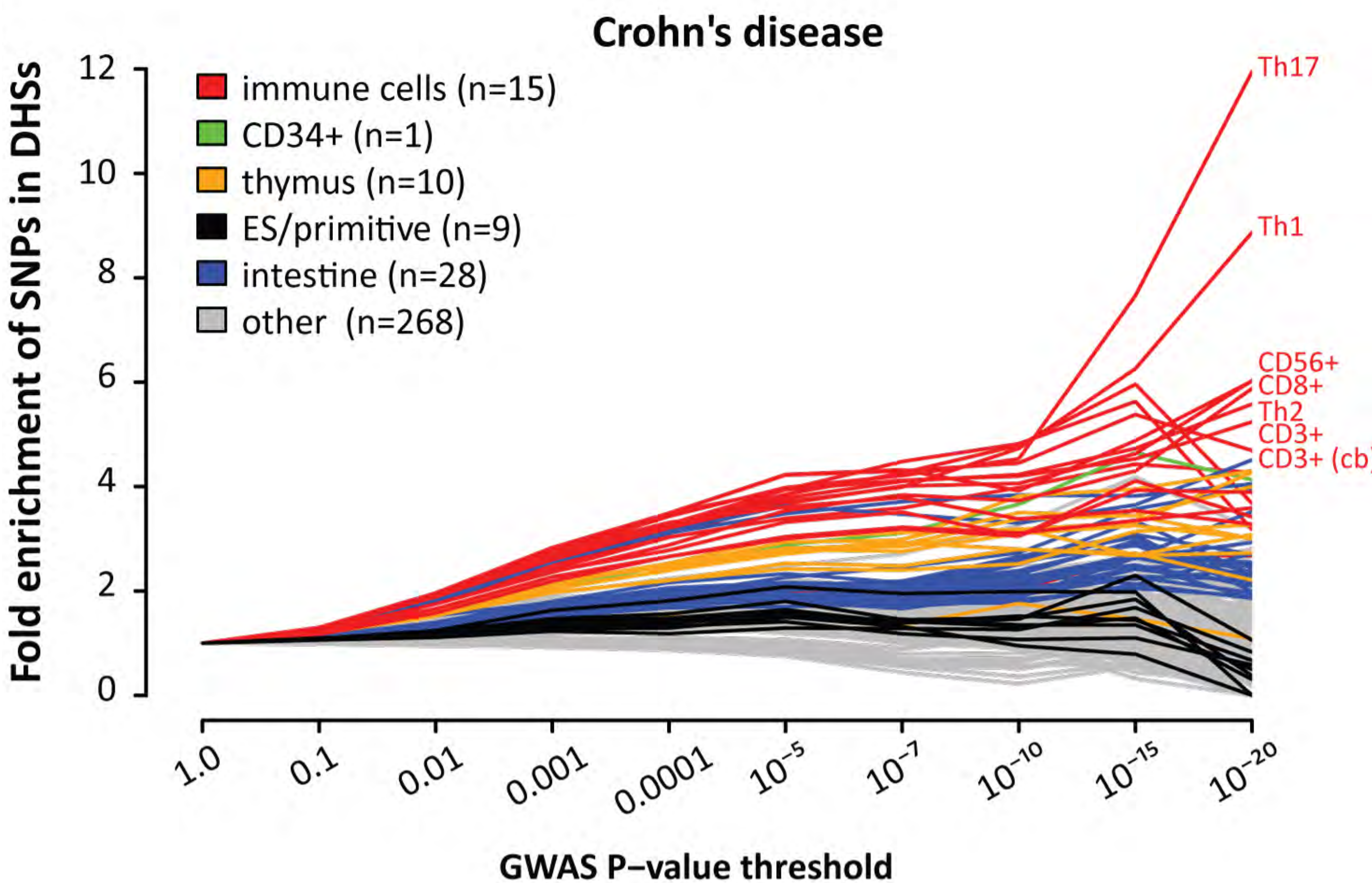
Maurano et al., Science 2012

Regulatory DNA maps enable pinpointing of disease/trait-relevant cell types

Cell-selective enrichment of trait-associated variants

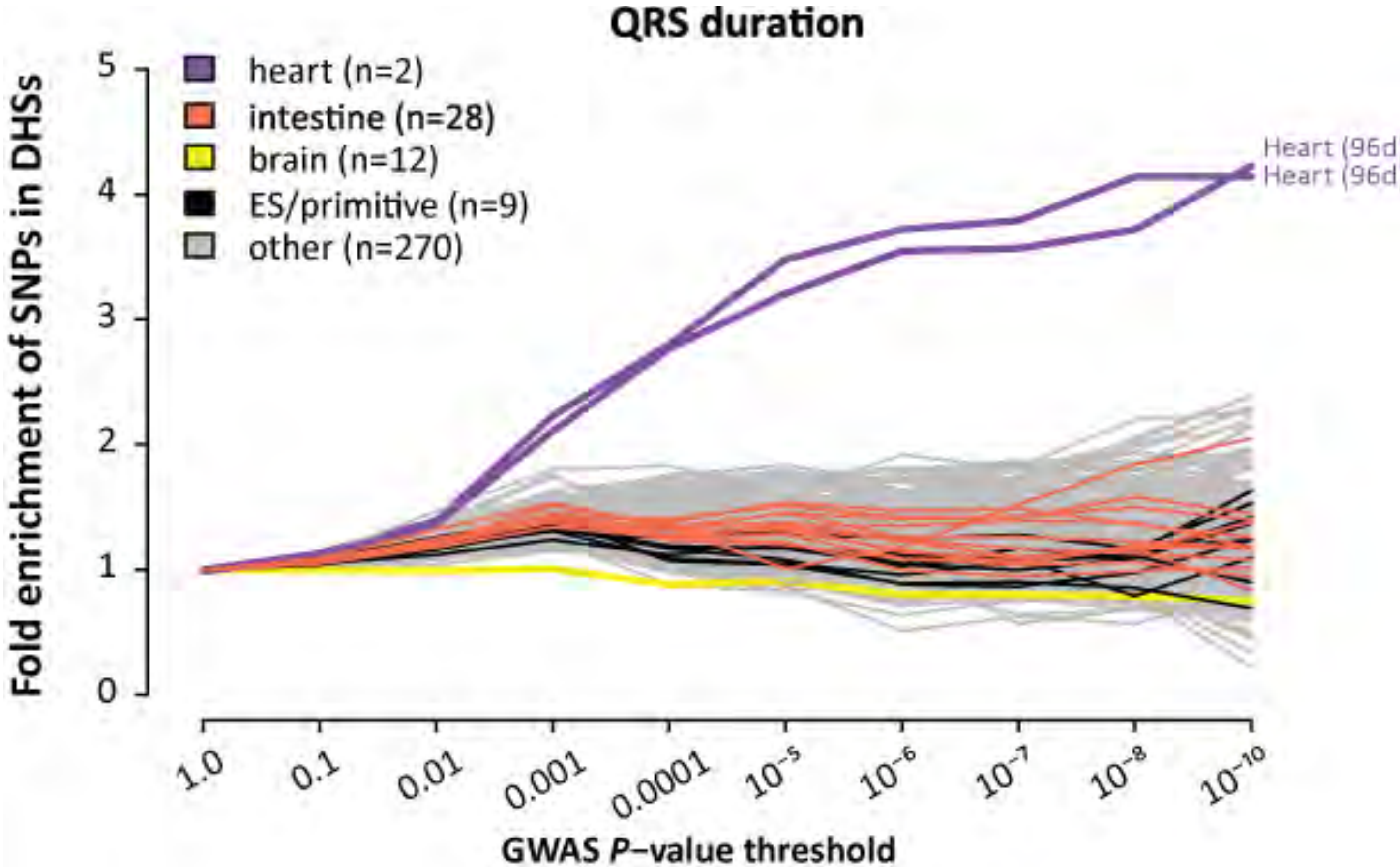


Selective enrichment of GWAS variants in pathogenic cell types



Maurano et al., Science 2012

Selective enrichment of GWAS variants in pathogenic cell types



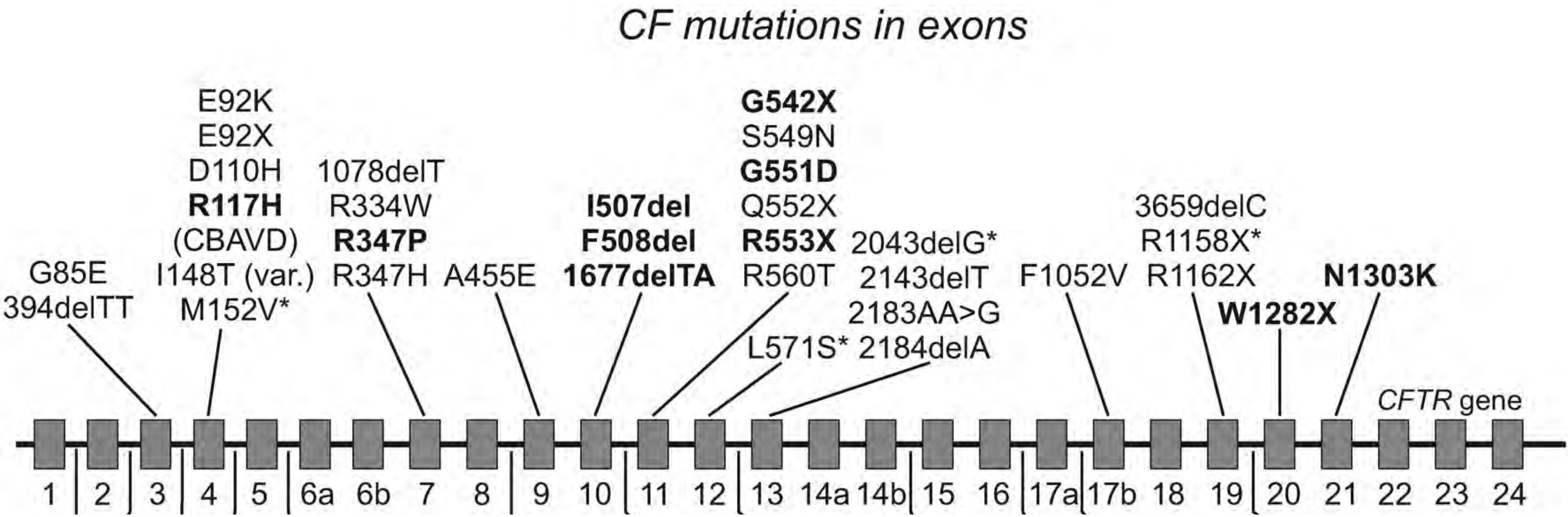
Maurano et al., Science 2012

Perspective:
Genes vs. ‘causal variants’

Many ways to break the clock

Functional regulatory vs. coding/splicing mutations

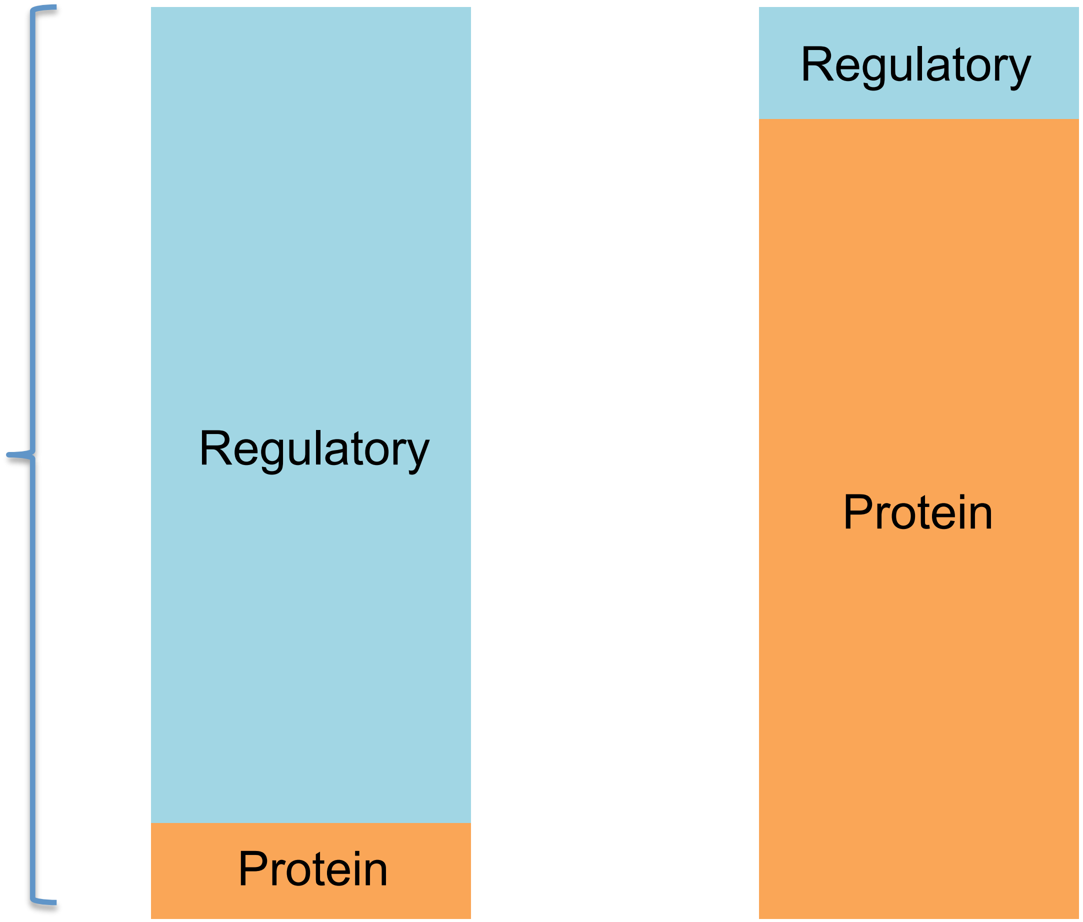
Why should we expect anything different?



Common Disease/ Trait

Rare (Mendelian) Disease/Trait

Nature & Proportion of variation



Effect of each variant

Very weak

Very strong

Acknowledgements

Key experiments/analyses
Matthew Maurano, Shane Neph, Jeff Vierstra
Tony Raubitschek, Andrew Stergachis,
Eric Haugen, Richard Sandstrom , Eric Rynes
Rich Humbert, Bob Thurman, Alex Reynolds

Epigenome Center
Raj Kaul, Scott Hansen, Peter Sabo, Molly Weaver, Theresa Cantwell, Kristin Lee, Shinny Vong, Vaughan Roach, Erica Gist, Sandra Stehling-Sun

Collaborators
Steve Ziegler (Benaroya Inst.)
Shelly Heimfeld (FHCRC)
Nona Sotoodehnia (UW Cardiology)
Chris Cotsapas (Yale)
Shamil Sunyaev (Harvard/Brigham)

Funding: NHGRI (ENCODE)
NIH Common Fund (Roadmap Epigenomics)